Vol. X Part II The British Journal of Statistical Psychology November 1957

PROFESSOR LEWIS M. TERMAN

Readers will have heard with deep regret of the death of Professor Terman. In this country he was perhaps best known for his revisions of the Binet–Simon scale. In the days when mental testing was still looked at askance by the majority of psychologists and educationists, he perhaps more than anyone succeeded in giving it a scientific status and putting it into practical shape. As he himself points out, he had no wish to contribute to statistical theory or to the principles of mental measurement, but only to make use of them. And by his own example he became one of the most effective advocates of metrical techniques, and did much to convince both teachers and administrators of the practical value of the new statistical procedures.

His origins, like that of so many eminent scientists, were curiously mixed. He was, as he himself puts it, " Scotch, Irish, Welsh, German, and French ". He was born in a country district about seventeen miles from Indianapolis on January 15, 1877. From data given by J. M. Cattell in *American Men of Science*, we learn that in a group of 1,000 scientists Terman was the only one belonging to a family of twelve or more. He was in fact the twelfth of fourteen children ; and his father was one of ten. The Termans, Tearmans, or Turmans were a roving and prolific family ; and his branch of it had apparently migrated Westward from Virginia. Most of them, like his grandfather and father, were hardworking and prosperous farmers. Nevertheless, the household library contained something like two hundred volumes : and the young Lewis apparently found books more interesting than the farmyard. Among his favourites he mentions Dickens, the *Encyclopaedia Britannica*, and *Peck's Bad Boy*.

There was, he says, nothing in his early environment that could have " conditioned him specifically in favour of psychology ". However, " almost as far back as I remember, I seem to have had rather more interest than the average child in the personalities of others, and to have been impressed by those who differed from the common run ". His elementary education was acquired in " a little red one-room school-house, which could not boast a single library book ". From eleven to eighteen he worked on the farm for about six months each year ; and " any mental development that occurred during this period must have been due to maturation rather than to intellectual stimulation ".

At school, and later on at college, he preferred linguistic subjects to mathematics and physical science. In this respect he notes a marked contrast between himself and another celebrated member of his class, Arthur Banta, the biologist. " Banta," he says, " was less bookish and less interested in the traits of people, but far more interested than I in plants, animals, and rocks." As it happened, both his school and his neighbourhood contained few boys of his own age. Most of them were older. And this, he thinks, may have provided ground for " the inferiority attitude and tendencies to introversion," which even in adult years he still revealed in the Bernreuter tests. A disposition of this type he

thinks may be somewhat characteristic of the psychologist, since it inclines a man towards introspection, and makes him secretly curious about the abilities and motives of those around him. Even before he reached his teens, he became fascinated by after-images, by the flight of colours, and by the connections between one idea and the next. He mentions an early discovery, which recalls a similar experience of Tennyson. " When about eleven or twelve years of age, I hit on the fact that, by repeating a formula, I could lose my sense of identity and my orientation in time and space : my method was to gaze fixedly on something, and repeat : ' Is this me ?' until all distinction between subjective and objective was lost in a kind of mystical haze."

While he was still only ten years old, " a pedlar who sold books on phrenology stopped at the farm for a night, discoursed on his science, and felt the bumps of each of us . . . When my turn came to be examined," says Terman, " he predicted great things of me ". That, he believes, may have stimulated him to greater efforts. And the interest in phrenology which was thus implanted served as his first introduction to the science of individual differences.

When he reached the age of fifteen, he went to a ' normal college ' to prepare for teaching. Of the various psychological textbooks that he studied, he " liked Sully " best. James' *Principles* he did not discover until later. But, he adds, " of all the founders of modern psychology my greatest admiration was for Galton ". At twenty-one he became principal of a township high school, where he " taught the entire curriculum to about forty pupils ". A year later he married. His wife relates how he once told her that his " interest in their baby had determined him to become a psychologist ". In 1901 he went to Indiana University ; and a couple of years afterwards to Clark University—in those days the Mecca for almost every aspiring young psychologist in America. He was one of a group of brilliant students, twelve of whom later became sufficiently distinguished to be included in the American *Who's Who.*

For a while he continued under " the hypnotic sway of Stanley Hall ". But he soon began to feel that the value of Hall's questionnaire procedures was getting exhausted. His own paramount interest still remained centred on individual differences, and more especially on methods of studying the more extreme cases—the gifted and the mentally deficient. He recalls two essays which he prepared for seminars—on ' The Great Man Theory ' and on ' Degeneracy '. Instead of the stock questionnaire, he began to consider the possibilities of mental tests. He thinks that, as with Thorndike, his mind may have turned in this direction partly because he " had no aptitude for, and little interest in, the laboratory techniques of the experimentalist of the Wundt-and-Titchener type ". Stanley Hall strongly disapproved. Terman, however, started to read whatever he could find on the subject, whether in English, French, or German : he mentions more particularly Galton, Binet, Stern, Cattell, and above all Thorndike. One thing, he found, was badly needed and that Clark University did not offer, namely, " instruction in statistical methods " : " a year with Thorndike would have been an untold boon ".

Unfortunately his studies were interrupted by a pulmonary haemorrhage and the discovery of a mildly active tuberculosis. On recovering, he was offered the chair of Child Study and Pedagogy at Los Angeles. Here he still had to nurse his health. But four years later in 1910 he was appointed professor of educational psychology at Stanford.

The first fruit of his lectures at Stanford was the small textbook on *The Hygiene of the School Child*. Huey, however, who with Gesell had for long been one of his closest friends, urged him to start an investigation on the value of the Binet scale of tests which had already been used for work on defectives by Goddard, Kuhlmann, and others. Accordingly, in collaboration with one of his graduate students, H. G. Childs, Terman began an experimental study of the 1911 scale which eventually led to the publication of *The Measurement of Intelligence* in 1916.

It was during the same period that Burt started his investigations on much the same problem ; and it is curious to observe how both of them, working to begin with, it would seem, with no knowledge of each other's interests, studied the same writers and adopted much the same lines of approach. Both Terman and Burt record their indebtedness to Galton's work in Britain, Meumann's in Germany, and Binet's in France. Owing, however, largely to the influence of Pearson, Burt became interested in the quantitative problems ; Terman, on the other hand, related that his " own interest in mental tests at that time was more in their qualitative than in their quantitative aspects ". But both appear to have reacted in much the same way to Spearman's methods and conclusions. Terman describes how, on reading Spearman's paper in the *American Journal of Psychology*, he was greatly impressed with " the dogmatic tone of the author, the finality with which he disposed of everyone else, and his one-hundred-per-cent faith in the verdict of his mathematical formulae "; but, he adds, " the conclusion to which it all led, namely, that there is ' a correspondence between what may provisionally be called General Discrimination and General Intelligence which works out with great approximation to *one* or *absoluteness* ', seemed to me as absurd then as it does now ".

The book on *The Intelligence of School Children* and the monograph on the Stanford Revision data were " by-products of the work that led to *The Measurement of Intelligence*". As in Britain so in America, " many of the old-line psychologists regarded the whole test movement with scorn "; " hence ", says Terman, he himself " was a little surprised that his publications were so favourably received ". Indeed, even at the time, he " imagined the revision would probably be displaced by something better ". And so eventually it was, namely, by ' the new revised Stanford–Binet tests ' produced by Terman in collaboration with Miss Merrill in 1937. British investigators who either then or later on were trying to modify or standardize the original Binet–Simon scheme, so as to render it more suitable for use in this country, will always remember with gratitude the generosity with which Terman permitted them to borrow or adapt either of his own revisions.

In the meantime, as soon as the first world war was over, Terman devoted his attention to the standardization of group tests of ability and scholastic tests for assessing educational achievements. At length a generous grant from the Commonwealth Fund enabled him to undertake the researches that he had always had in mind on ' the psychology of gifted individuals '. The result was the three large volumes, published between 1924 and 1930, under the title of *Genetic Studies of Genius*. A few years later, he turned to what at first sight appears a somewhat unexpected field of investigation, although, as he himself records, the problem had been germinating in his mind for many years as a result of his comparative studies of boys and girls during their years of growth and schooling. The outcome was the work on *Sex and Personality*, based largely on his well known ' Masculinity-and-Femininity Index ', and published in collaboration with C. C. Miles.

In his last publication, his Walter Bingham Lecture (1954), he reverted once again to the topic nearest to his heart, ' The Discovery and Encouragement of Exceptional Talent '. The problem of the gifted individual, he maintains, still fails to receive the attention it merits ; and for this reason he is " inclined to be pessimistic about present trends in democracy ". His closing words are well worth quoting. " What I especially want to emphasize," he says, " is that the evidence on early mental development of historical geniuses and the follow-up of gifted subjects selected by mental tests point to the conclusion that capacity to achieve beyond the average can be detected early in life by a well-constructed ability test heavily weighted with the *g* factor."

It would be presumptuous to attempt any review of Terman's great and lasting contributions to psychology. But he himself has summarized the salient points. " If I am remembered after my death," he writes, " it will probably be in connection with my studies of gifted children, the construction of mental tests, and the psychology of sex differences. I think that I early saw, more clearly than others, the possibilities of mentality testing, and have succeeded in devising tests that work better than their competitors, and, by the application of test methods, have added to the world's knowledge of exceptional children." He believes that " the revival of associationism by ' bond ' psychology has run its course : the Watsonian brand of behaviourism is a cult ; and a denial of heredity on the basis of a few minor experiments on the conditioning of infants ridiculous ; the major differences between children of high and low I.Q., and the major differences in the intelligence test scores of certain races, will never be fully accounted for on the environmental hypothesis." These conclusions plainly have an importance beyond the bounds of psychology; and, if they are gradually winning acceptance, that is largely due to the long and tireless labours of Professor Terman himself.

WILLIAM B. LEWIS.

Vol. X
Part II

The British Journal of Statistical Psychology

November
1957

THE RELATIONS OF THE NEWER MULTIVARIATE STATISTICAL METHODS TO FACTOR ANALYSIS

By Harold Hotelling
University of North Carolina

Abstract. A survey of developments in multivariate analysis during the last thirty years shows that some, though not all, of the purposes for which factor analysis has been used may now be better accomplished by other procedures, e.g. by regression, multiple correlation, the study of relations between two sets of variates, or of the dimensionality of a set of variates.

To determine whether two or more groups of persons differ significantly in their mean values or their covariance matrices, the most appropriate procedures consist of methods of multivariate analysis of variance. Such methods have increased rather than diminished the advantages of using an external criterion instead of making a purely internal analysis.

The problem of estimating the dimensionality of a continuous multivariate population by means of a sample admits a simple and exact answer : the rank of the population is equal to that of the sample, provided the number of degrees of freedom among the individuals in the sample exceeds the number of the variates. If, however, the problem is re-interpreted so as to imply that each observation is the sum of a real part and a random error, then the hypothesis that the real part has a rank smaller than the number of variates can be tested only when a proper estimate of the errors is available, i.e. by making suitable replications. The appropriate method is again multivariate analysis of variance rather than factor analysis. In a third form of the problem the object is, not to ascertain a true dimensionality, but to assess the error resulting from discarding (on practical grounds) a dimension that may really exist. Here the problem is somewhat indeterminate : a practical expedient may involve a special form of factor analysis, viz. the calculation of the least, not the greatest, latent roots.

Thus, in many cases in which they have hitherto been used, factor analyses of the usual kinds are inferior to other procedures : nevertheless, the results of such analyses may have heuristic and suggestive value, and may uncover hypotheses which are capable of more objective testing by other methods.

I. Recent Developments in Multivariate Analysis

Factor analysis in psychology, and the more general study of multivariate analysis in statistics, have developed side by side and to a great extent independently. Owing to lack of adequate communication between the two groups of research workers—each spread geographically, but having few contacts with the other—psychologists have often employed statistical methods demonstrably inferior to those worked out by mathematical statisticians, while the

latter have often been unaware of the highly interesting activities and stimulating problems of psychologists. The following paper outlines achievements in multivariate statistical analysis during the last quarter-century, suggests uses in psychology of the newly discovered methods, and tries to mark out a boundary between the cases in which these should supersede methods involving factor analysis and those in which factor analysis is still the most useful tool. It will be seen that some, but not all, of the things people have undertaken to do by factor analysis can be done better in other ways. The methods here considered have all been derived primarily for use with multivariate normal distributions; in other cases they still have certain descriptive, suggestive, and approximative values.

(i) A primary tenet of modern statistics is that inferences must be supported by known probabilities to keep under control the effects of errors in measurement and in sampling. (ii) In selecting statistical methods a second guide-post is that they must maximize the probabilities of right conclusions and minimize the probabilities of wrong. This position, which was illustrated by Gauss' espousal of the mean square error as against less efficient measures of dispersion, has been strongly emphasized in a series of influential papers by R. A. Fisher, beginning with his memoir on 'The mathematical foundations of theoretical statistics' [5]. It implies that the *choice* of methods of descriptive statistics is logically posterior to the *study* of statistical inference, though the *practice* of descriptive statistics comes normally before the inference. (iii) Thirdly, the statistician has abandoned his old passive role of merely analysing whatever was put on his doorstep by other investigators or by chance, and now has much to say about how experiments should be designed to make inferences from them valid and the whole process economical.

The first principle, calling for known probabilities, has stimulated much ingenious research in methods of finding probabilities; but it has also done something more fundamental in stipulating that the probabilities are to be known numbers wherever possible, and not merely functions of unknown parameters that can only be estimated. It was to satisfy this stipulation that the 'Student' distribution was introduced, and later on, the analysis of variance and the exact distributions of the ordinary correlation coefficient, the multiple and partial correlation coefficients, the multivariate generalization of the Student distribution, the multivariate forms of analysis of variance, and canonical correlations. Here again it was R. A. Fisher who obtained the first of these exact distributions, and led the way in showing how to put them to practical use.

The multivariate normal distribution in p variates, even with zero means, has $p(p+1)/2$ independent parameters; and the distribution of an arbitrarily chosen 'statistic', or function of the observations, may involve all of them. If p is at all large, the number of these parameters is very large indeed. When, as is usual, none of them is known apart from the observations, the control of error probabilities becomes hopelessly difficult unless the statistics to be used are carefully chosen so as to keep down the number of these unknown quantities,

which I have termed 'nuisance parameters', entering into the most relevant probabilities.

This need of keeping down the nuisance parameters limits drastically the statistical methods appropriate for use with multivariate normal distributions, going beyond the second, or 'efficiency' tenet, which leads to the use of sufficient statistics where they exist, as they do the multivariate normal distribution. The sufficiency criterion tells us here that we should use no statistic that is not a function of the sample means, variances, and covariances. The new criterion tells us further that among these functions we should seek to use only those in whose distributions the nuisance parameters can be kept under control and ideally are non-existent. Many of the powerful methods discovered in the last three decades are based on this principle.

The first example is the multiple correlation coefficient, whose general distribution in samples from a multivariate normal population was discovered by Fisher [6]. It looked at first as if the distribution would involve all the covariances as independent parameters, and therefore be excessively complicated and difficult to determine as a function; moreover, even if the distribution function of R were determined, it would be useless if it involved all these many nuisance parameters, since they are themselves unknown. Fortunately the situation was saved by the invariance of R under all non-singular linear transformations of the predictors, or 'independent variables'. Such a transformation can always be found that will make the predictors mutually uncorrelated, and will also make all but one of them uncorrelated with the predictand. Thus in the transformed system there is only one correlation that can be different from zero, and this equals the population multiple correlation coefficient ρ. Hence the distribution of R involves only ρ; and Fisher, using this invariance property and some ingenious geometry, with other tricks, found the exact probability density.

The great principle illustrated by Fisher's success with the multiple correlation coefficient is the possibility of eliminating nuisance parameters and simplifying exact probabilities when the statistics used are invariant under all non-singular linear transformations of the variables, or even under an extensive class of such transformations such as the orthogonal. This puts a premium on the use of invariant statistics, and points to the selection of such statistics to meet various needs. The criterion of invariance has been influential ever since Fisher's 1928 paper in the development of new methods of multivariate analysis.

II. The Generalization of Student's Distribution

The manifold uses of the Student distribution in one dimension, for example in testing the significance of means and differences of means, and in laying down confidence intervals with known probabilities, had long required generalization to two or more dimensions, and various unsuccessful attempts had been made in this direction. Notable was the 'Coefficient of Racial Likeness', proposed by Karl-Pearson [13], which, like other such efforts, foundered on the nuisance

parameter rocks. It was the sum of squares of differences between corresponding means in two samples, with each square divided by a variance estimate drawn from the data of both samples. Each term represented some measured trait. The trouble was that this statistic, which was used with enthusiasm in the 1920's, had a distribution that was not only impossibly hard to find, but involved all the unknown correlations and ratios of true variances among the variates.

The remedy eventually found for this trouble was the replacement of the simple diagonal matrix of the quadratic form in the differences of the means by a matrix inverse to that of the sample covariances among these differences. The square root T of this quadratic form is a direct generalization of the t of the Student distribution, and for p variates has all the uses that t has for one. For example, it may be desired on the basis of a random sample of N individuals from a p-variate normal distribution to define a region in a space of p dimensions which will have a probability 0·95 of covering the point whose coordinates ξ_i are the true but unknown mean values of the variates in the population. This may be accomplished as follows. Defining $n = N-1$, $s_{ij} = S(x_i-\bar{x}_i)(x_j-\bar{x}_j)/n$, where the summation is over the sample, we invert the p-rowed square matrix $[s_{ij}]$ to get $[l_{ij}]$, and consider the quadratic form

$$T^2 = N \sum\sum l_{ij}(\bar{x}_i-\xi_i)(\bar{x}_j-\xi_j).$$

The value of this expression is invariant under all non-singular linear transformations among $x_1, \ldots, x_p$; and this fact makes it possible to obtain its distribution exactly and in a relatively simple form, easily transformable to the incomplete beta function or variance ratio distribution. Since tables of the latter are now widely available, it is convenient to obtain from them the numerical value F having probability 0·05 of being exceeded, with $n_1 = p$ and $n_2 = n-p+1$ degrees of freedom, and then to compute

$$T^2 = npF/(n-p+1).$$

Equating this value of T^2 to the quadratic expression above gives the equation of the boundary of the required confidence region in the space of $\xi_1, \ldots, \xi_p$. I have given the full derivation, valid both for this and for various other uses, in the *Annals of Mathematical Statistics* [8], and have discussed and illustrated these methods further in *Statistics and Mathematics in Biology* [11]. The general formula for T^2 is

$$T^2 = \sum\sum l_{ij}\delta_i\delta_j,$$

where the δ's have the same covariance matrix as the x's and zero means, and are independent of the sample covariances. Thus in testing the significance of the differences between the means in two independent p-variate random samples from normal distributions with the same covariance matrix, we put

$$\delta_i = \frac{\bar{x}_i - \bar{x}'_i}{\sqrt{\left(\frac{1}{N}+\frac{1}{N'}\right)}},$$

where the sample numbers are N and N'. The distribution is the same as before, with $n = N+N'-2$; this n is also the denominator of each of the covariance estimates, which are formed by pooling corresponding product sums in the two samples.

S. S. Wilks [17] studied certain generalizations of the analysis of variance to p dimensions, using determinants and ratios of determinants having invariantive properties under linear transformations. Other types of multivariate analysis of variance have since been investigated by S. N. Roy, P. L. Hsu, M. S. Bartlett and others. These are suitable, among other things, for testing significance of differences among three or more groups, representing for example different treatments, and generalize the use of T as the ordinary univariate analysis of variance generalizes the use of the Student t.

If in the formula above for T^2 in terms of the δ_i the latter do not all have zero expectations, the distribution is of a more complicated form which was obtained by R. C. Bose and S. N. Roy [4]. As in the case of the multiple correlation coefficient, the invariantive property made it possible to show that only one parameter enters into the distribution, and enabled Bose and Roy to obtain the probability element as a multiple of a confluent hypergeometric function. When the δ_i are suitable multiples of the differences between corresponding sample means in two populations, the parameter Δ^2 in the distribution of T is a quadratic form of the same type as T^2, with population instead of sample means and covariances; Δ may then be called the figurative distance between the two populations if these have the same covariance matrix, and possesses the properties of distance, including the triangular inequality. The Pearson Coefficient of Racial Likeness should be replaced by T^2 for samples, by Δ^2 for populations. Much empirical work using this modernized coefficient of racial likeness has been carried out in Bengal under the leadership of P. C. Mahalanobis. For unequal covariance matrices in the two populations and matched samples, T may be used in an obvious way generalizing Student's celebrated method of comparing hours of sleep gained by using two isomers of hyoscyamine hydrobromide.

III. The Relations between Two Sets of Variates

A different kind of invariance is involved in the determination of relations between two sets of variates, as in my paper in *Biometrika* [9]. One set might consist for example of scores in several mental tests given to army recruits, and the other of scores on various kinds of subsequent performance. The determination of a linear function of the test scores and a linear function of the later performance scores, such that the correlation between these two linear functions is as great as possible, leads to a determinantal equation whose greatest root is the required maximum correlation. The other roots determine other matching pairs of linear functions, such that each of these linear functions of the variates in one set is uncorrelated with all but its own partner in the other set. These functions are the canonical variates; and the correlations between matching pairs

are the canonical correlations between the two sets of variates. If the canonical correlations are all zero, the two sets are completely independent; and it is useless to predict performance by means of any of the tests used. If one of the canonical correlations is unity, there is something about the performance, represented by a certain linear function of the performance measures used, that can be predicted perfectly by means of the tests. If two canonical correlations are unity, there is a whole one-parameter family of types of performance that can be predicted perfectly by means of these tests.

A use sometimes made for factor analysis in the past is in testing for the relations between two sets of variates. One study of the relations of character with mentality, for example, used seven mental tests and seven estimates of character traits by acquaintances. A standard factor analysis technique was applied to the mental measurements, and independently to the character scores, so as to get seven mental factors and seven character factors. A name was applied to each of the fourteen factors thus found; and a plausible matching of character factor with mental factor was arranged so as to get seven pairs. The correlations within the pairs thus calculated were then computed, and judged in each case insignificant.

This kind of use of factor analysis should clearly be superseded by an examination of canonical correlations between the two sets of variates. The largest canonical correlation will be larger than all the correlations between the two sets actually found, and has a better chance of being found significant and thus demonstrating the existence of a relation between the two sets if one actually exists. The other canonical correlations, in decreasing order of size, and the corresponding canonical variates, will help to elucidate the nature of any relations that may exist between character and mentality. The separate factor analyses of the two sets were really quite useless for this purpose.

A simpler illustration of the need at times to resist the temptation to perform factor analyses is provided by ordinary regression problems, in which a least-squares formula for a single variate y is to be determined by variates $x_1, \ldots, x_p$. This is indeed a special case of two sets of related variates, in which the number in one set reduces to unity, the maximum canonical correlation to the ordinary multiple correlation, and the other canonical correlations to zero. If as a preliminary one performs a complete factor analysis on $x_1, \ldots, x_p$—let us say a resolution into principal components—the resulting 'factors', or components of the complex will have exactly the same multiple correlation with y as did the original x's; and a regression equation for y with them as predictors will for every individual give the same prediction as if the factor analysis had not been performed. The labour of solving the normal equations will be greatly reduced if orthogonal factors are used, but this saving is trifling as compared with the heavy labour of the factor analysis itself if p is large. (Here of course exceptions must be made, including the possibility of electronic computers that will reduce all this labour virtually to zero; also the remote chance of all the correlations among the original variates being zero.) But if not all the principal components are used

in the regression analysis, the results will be virtually affected. If even the last principal component, the one making the least contribution to the total variance, is omitted, both the regression function and the multiple correlation coefficient may be drastically different.

To see this, consider the case of two predictors, x_1 and x_2, which with the predictand y are represented by three vectors through the origin in the sample space. The multiple correlation is the cosine of the angle θ made by the y-vector with the plane of the two x-vectors, and the regression values are represented by the vector **Y** within this plane making a minimum angle with y, and therefore the orthogonal projection of y on the plane. Replacement of the two predictors by their two principal components or other 'factors' is equivalent to rotating each of the x-vectors within their common plane, a process that does not affect the regression vector **Y** nor the angle θ, since these depend only on the plane and on the external predictand (or criterion) y. But suppose only one of these 'factors' is used. The plane now collapses into one of the lines in it; and the regression vector is then forced to lie along this line, which is not determined by y, but only by the x's. There is nothing to prevent, or even to make improbable, a great increase in the angle of the new regression vector with y, with a corresponding decrease of the correlation and of the accuracy of prediction. It is possible for even a high correlation to be reduced to zero when the set of predictors is replaced by a reduced number of linear functions of them, such as might be obtained by factor analysis, and in virtually every case there will be some decrease of correlation and of the accuracy of the prediction formula obtained.

The successful determination of exact probability distributions for the multiple correlation, canonical correlations, the generalized analysis of variance and Student ratio, and certain other modern methods has in each case hinged on the invariantive property of the statistics used. It is apparently only by restricting the choice of statistics to such invariants that the appalling number of nuisance parameters in the distributions of functions of numerous correlated variates can be reduced sufficiently to bring the sampling errors under control.

The necessity of sampling distributions in drawing any reliable inferences thus points to the use only of functions of observations possessing invariantive properties, since no exact sampling distributions seem to be obtainable otherwise. It is true that asymptotic standard errors are used in default of exact distributions; but in the absence of any known bound for the error resulting from this practice, it cannot be recommended excepting in desperate extremities. Far better is the use of statistics for which exact distributions can be determined, 'exact' in the rather technical sense of freedom from any unnecessary dependence on nuisance parameters, and 'exact' also in the ordinary sense of accurate computability.

In a matrix of covariances the only thing invariant under all non-singular linear transformations of the variates is the rank. For covariance matrices of observations, constituting a sample from a multivariate normal or any other continuous distribution, the actual rank is virtually always equal to the number

of variates, and therefore has no real interest. In order to achieve invariant statistics of real value in this connection we need either *two* matrices relevant to our problem, or some relevant restriction of the class of transformations.

In the latter category is a restriction to orthogonal transformations. Under this subclass of all linear transformations there are invariants, the latent roots and latent vectors of the matrix, corresponding geometrically to the principal axes of an ellipsoid determined by the matrix. This is one way in which we are led to the method of principal components. There are also other ways, including several minimizing or maximizing properties. But one may wonder why orthogonal transformations should be given such a position of pre-eminence. A change of units of measure, for example, is a non-orthogonal transformation, but may often be desirable, and is actually made when the standard deviation of each variate is taken as its unit of measure, thus replacing the covariances by correlations.

When we come to the comparison of two covariance matrices, the number of invariants under all non-singular linear transformations is significantly increased. For a pair of symmetric square p-rowed matrices A and B, of which B at least is non-singular, a complete set of invariants is provided by the roots $\lambda_1, \ldots, \lambda_p$ of the determinantal equation

$$|A - \lambda B| = 0,$$

each with an associated vector. No restriction to orthogonal transformations is now needed to get meaningful invariants.

If A and B are independent sample covariance matrices corresponding to the same population covariance matrix, the joint distribution of $\lambda_1, \ldots, \lambda_p$ involves only these statistics themselves, the number of variates p, and the numbers of degrees of freedom of the two sets of sample covariances. This distribution was determined virtually simultaneously by S. N. Roy [14], [15], P. L. Hsu [12], and R. A. Fisher [7]. To test the hypothesis that two independent sample covariance matrices for the same p variates have the same expectation, one may select some function of $\lambda_1, \ldots, \lambda_p$ that will be sensitive to a particular type of deviation from equality against which the null hypothesis is to be tested. There is then the problem of deriving the distribution of the particular function chosen from the Roy–Hsu–Fisher distribution, as by transformation and integration to pass from joint marginal distribution. This process has been carried out exactly for several cases for which $p = 2$, and approximately in other cases.

It is in this way that the several forms of multivariate analysis of variance may be regarded. The matrices A and B may be obtained from independent samples, or where there is replication, from the same sample, by obvious extension of the separation in univariate analysis of variance of the error sum of squares from the rest. The familiar univariate ratio F of the sum of squares of main concern to the error sum of squares, multiplied by a ratio of numbers of degrees of freedom, is generalized, if B is the matrix of error covariance estimates, by either AB^{-1} or $B^{-1}A$. Each of these matrices must be represented

by one number in order to carry out a test of significance to see whether A represents deviations substantially beyond what can be expected by chance of the sort revealed by the discrepancies between replications summarized in B. This number may be the determinant

$$|AB^{-1}| = |B^{-1}A| = |A|/|B| = \lambda_1, \lambda_2, \ldots, \lambda_p,$$

whose distribution was studied, with this use in view, by S. S. Wilks [17]. It is the statistic indicated by the likelihood ratio criterion. The use of the greatest of the roots has been developed by M. S. Bartlett [1] and later papers [2, 3], and S. N. Roy in the multivariate analysis of variance [16], and presents advantages in numerous applications.

Another method makes use of the sum of the roots. This arose originally from use of the generalized Student ratio T^2 to measure the deviation of a bomb from its target, against a background of recorded errors in range and deflection of such bombs. I have written about this distribution in three symposium volumes: uses of it in *Techniques of Statistical Analysis* (1947), its mathematics in the *Proceedings of the Second Berkeley Symposium* [10], and its relation to other multivariate methods in *Statistics and Mathematics in Biology* [11]. The joint distribution of the squares of the canonical correlations is closely related to that of $\lambda_1, \ldots, \lambda_p$, and has been studied by P. L. Hsu in *Biometrika* for 1941.

Problems of discrimination and classification among populations and individuals have been considered by many writers. Factor analyses appear to have little place in this problem, though it is occasionally mentioned. A. Wald proposed (*Annals of Mathematical Statistics*, 1944) a statistic for measuring the extent of misclassification, and with the help of its invariantive properties reduced its distribution to a triple integral. This was studied further by H. L. Harter (*Ann. Math. Stat.*, 1951).

IV. The Dimensionality of a Multivariate Normal Distribution

Let us consider one further problem of long standing that helps to bring out the relations between factor analysis and methods of multivariate analysis of recent development. This is the question of the dimensionality of a multivariate normal distribution. The dimensionality is the same as the rank of the matrix of the observations, the covariance matrix, and the correlation matrix. Frequently asked questions are how, on the basis of a sample, to determine this rank, and how accurate the determination is. The answers are, in one sense, easy. The rank of a sample cannot be greater than that of the population from which it is drawn. For example, if a distribution of points in three-dimensional space is confined to a plane, then every sample of these points lies in the same plane. On the other hand, for a distribution having continuous positive density of dimensionality p, the probability that a sample of more than p points should lie entirely within a flat space of dimensionality less than p is zero. If the sample consists of p or fewer points, the dimensionality is one less than the sample size. Thus, apart from possibilities whose total probability is zero, we may say that

if there are more individuals than variates, the rank of the sample matrices is exactly that of the population. This is a complete answer to the questions as they are most often asked.

This answer is however unlikely to satisfy our questioner. The rank of an empirical matrix of observations (or of their correlations or covariances) is practically always equal to the number of variates; and he wants to represent these by a smaller number of variates, and is looking for a statistician's blessing on this labour-saving procedure. His real problem is not whether in reality there are fewer dimensions than variates, but whether he will make much of an error if he proceeds as if there were. It is of course impossible to give him satisfactory assurances without knowing exactly what it is that he is ultimately going to do with his calculations, and even if one does know this it is difficult. There is a chance of considerable mischief ultimately developing from neglect of something that is really there, but since so much has to be neglected, it is not seriously out of character to suggest, for example, the method of principal components to pick out a small number of 'factors' to represent the whole lot. This can be done without any attempt to judge the 'reality' of the factors.

Another interpretation is worth noticing. Perhaps each observation is the sum of a real part and a random error, and the real parts have fewer than p dimensions, though the observations are of rank p. This is a self-consistent hypothesis. To test its correctness we need one of two things. Either we must know the matrix of covariance *of the errors*, or we must be able to estimate this matrix from observations. The latter requires replication, so that independent estimates shall be available of the same true part. If q independent estimates, subject to unbiased errors having the same covariance matrix, are available for each individual, a multivariate analysis of variance may be employed to test the significance of the deviations from any particular hypothetical dimensionality. This procedure is capable of providing definite tests in terms of probability, and does not require a factor analysis.

Factor analysis has often been thought of as a means of finding genetic factors; indeed, this seems to be the origin of the word 'factor' in its name. It would be more in keeping with geometrical and physical traditions to speak of 'components' of a tensor, which would be 'principal components' if possessed of certain extremal properties. The determination of genetic factors requires comparison of individuals carefully classified according to their relationships. A matrix of correlations referring to persons of unknown degrees of relationship cannot possibly prove or disprove anything about genetic factors, whatever mathematical procedures may be applied to the matrix. Thus the scope of valid applications of factor analysis in the usual sense must be reduced by the elimination of genetics.

Acknowledgments

The substance of the foregoing paper was discussed at the recent International Colloquium on Factor Analysis, and a French translation published in full in the Report of the Proceedings (Paris, 1955). The author is indebted to the authorities

of the *Centre National de la Recherche Scientifique* for permission to publish an English version here.

References

[1] Bartlett, M. S. (1934). Probability and chance in the theory of statistics. *Proc. roy. Soc.* A, CXLI, 518.

[2] Bartlett, M. S. (1948). Internal and external factor analysis. *Brit. J. stat. Psychol.*, I, 73.

[3] Bartlett, M. S. (1950). Tests of significance in factor analysis. *Brit. J. stat. Psychol.*, III, 77.

[4] Bose, R. C., and Roy, S. N. (1938). The distribution of the studentised D^2-statistic. *Sankyā*, IV, 19.

[5] Fisher, R. A. (1921). On the mathematical foundation of theoretical statistics. *Phil. Trans. roy. Soc.* A, CXXII, 309.

[6] Fisher, R. A. (1928). The general sampling distribution of the multiple correlation coefficient. *Proc. roy. Soc.* A, CXXI, 654.

[7] Fisher, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen.*, IX, 238.

[8] Hotelling, H. (1931). The generalization of 'Student's' ratio. *Ann. math. Statist.*, II, 360.

[9] Hotelling, H. (1936). Relations between two sets of variates. *Biometrika*, XXVIII, 321.

[10] Hotelling, H. (1950). Ap. *Proceedings of the Second Berkeley Symposium.* University California Press.

[11] Hotelling, H. (1954). Ap. *Statistics and Mathematics in Biology.* Iowa State College Press.

[12] Hsu, P. L. (1939). On the distribution of roots of certain determinantal equations. *Ann. Eugen.*, IX, 250.

[13] Pearson, Karl (1926). On the coefficient of racial likeness. *Biometrika*, XVIII, 105.

[14] Roy, S. N. (1939 a). A note on the distribution of the studentised D^2-statistic. *Sankyā*, IV, 373.

[15] Roy, S. N. (1939 b). p-statistics, or some generalisations on analysis of variance appropriate to multivariate poblems. *Sankyā*, IV, 381.

[16] Roy, S. N. (1942). Analysis of variance for multivariate normal populations. *Sankyā*, VI, 35.

[17] Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika* XXIV, 471.

Vol. X
Part II

The British Journal of Statistical Psychology

November
1957

THE DISTINCTION BETWEEN COMMON AND SPECIFIC VARIANCE IN FACTOR THEORY

By Charles Wrigley[1]
University of California

Abstract. The purpose of this paper is to indicate certain inadequacies in the current use of ' communalities ' in factor theory. There are considerable divergences in the way the term is defined ; and several reasons for believing that, as ordinarily computed, the values become highly unstable when the number of tests or the size of the sample is increased. It is accordingly suggested that a better measure of the common factor variance might be furnished by substituting the square of the multiple correlation.

I. The Common Factor Theory

Factor theory assumes the variance of any test to be divisible into three parts : (i) variance which is common to other tests as well; (ii) variance which is reliable but specific to that test; (iii) error variance which arises from test unreliability. Specific and error variances may be combined to form the unique variance. This threefold distinction was introduced because of the unusual limitations shown by most psychological measurements. The psychologist naturally wants to exclude from his analysis error variance of the kind revealed by changes in scores from one occasion to another, or from parallel test forms used on the same occasion; and the reliability coefficient might seem the appropriate value to insert in the diagonal in order to achieve this. But he also seeks to exclude specific variance from his analysis, since, unless reliable variance can be shown to be present in two or more tests, it may conceivably represent something peculiar to one particular test only. These two contributions to the total variance—the specific and error variance—not only introduce further factors of no psychological interest, but may even become confounded with the common variance, thereby confusing the description of factors having psychological relevance. Hence it is commonly argued that the factorist should restrict his analysis to the common factor variance.

The term ' communality ' has been defined by various writers in different and sometimes conflicting ways. Through all their definitions, however, there runs the notion that communalities are reduced values to be inserted in the leading diagonal, and that, after extraction of k common factors ($k < p$, the number of tests), they become either zero or very small. The communalities are taken to represent the ' common variance ', and the variance which is thus excluded from the diagonal entry to represent the ' uniqueness '.

[1]This research was supported by the United States Air Force under Contract No. AF 41 (657)—76, monitored by Air Force Personnel and Training Research Center, ATTN : Dr. John M. Leiman, Personnel Research Laboratory, Post Office Box 1557, Lackland Air Force Base, San Antonio, Texas.

The disadvantages of introducing these reduced entries are well known. Adoption of a factorial structure with more factors than tests complicates the algebraic exposition and the difficulties of calculation. There is no generally accepted method for finding the communalities; and all take considerable time for a computer with only a desk machine. Yet most factorists regard these difficulties as a reasonable price to pay for reducing the lack of precision.

In factor analysis there is always a twofold sampling. A sample of persons is selected from a population of persons and a sample of tests from a population of tests. To avoid confusion, the population of tests will be called the *domain* (Tryon's convenient term [30], p. 4), and the sample of tests the *selection*. Alternatively, Guttman's term *universe* ([15], p. 282) might be used for the population of tests, and *battery* could be used for the sample. The 'squared multiple correlation' between each test and the remaining $(p-1)$ tests will be referred to as s.m.c.

II. Inadequacies of the Common Factor Theory

The theory of common factors has four main inadequacies : (i) the existence of alternative sets of communalities satisfying the theory; (ii) the relativity of the distinction between common and specific variance, which depends in part on the particular selection of tests; (iii) the failure to distinguish explicitly between common and significant factors; (iv) when significant factors are insisted upon, the possibility of changes in the probability level adopted for significance or of an increase in sample-size resulting in more factors and increased values for the communalities. In certain forms of analysis there is a further difficulty, which is not a necessary consequence of the theory, but interlocks with it. Factors are usually rotated. If their number is increased by increasing the size of the sample, a restructuring of rotated factors is likely, with the larger factors splitting into smaller.

(i) *Alternative Sets of Communalities.* Various sets of communalities, based on varying numbers of common factors, will satisfy the theory, i.e. provide residual correlations of exactly zero [22]. The solution with the fewest common factors is generally assumed to be the best. To those who hold that factors are only operational fictions, convenient for classifying tests, this principle of parsimony presents no special difficulty. Since the s.m.c. of any test measures only common variance, a satisfactory case can be made for accepting the set of communalities with minimal deviations from the s.m.c.'s, so long as we are content with operational definitions. But to those who consider factors to represent real entities, the matter is not so easy, since it is hard to see any strong psychological reason why the solution with fewest factors is more likely to be the 'true' representation. This difficulty becomes still more acute when various sets of communalities are possible for the same minimal number of factors : see Thomson ([26], pp. 81–82) for an example.

(ii) *Relativity of the Distinction.* The distinction between common and specific variance is relative, and depends on the particular tests selected. By

adding suitable tests, specific variance can be converted to common variance. In Thurstone's analysis of perceptual tests ([28], p. 96), the white-black flicker fusion test is almost completely specific, with a communality of only 0·12. But elsewhere in the same study (p. 23), a correlation of 0·41 of the white-black flicker fusion test is reported with the blue-yellow flicker fusion test (a test excluded from the factor analysis). Presumably a flicker fusion factor would have been isolated if tests for white-black, blue-yellow, and red-green flicker fusion had all been included; and the communality for the white-black test would then have been much higher.

On the other hand, common factor variance can be reduced to specific by excluding all but one of the tests with appreciable loadings for the common factor. Thurstone's perceptual study again provides an example. The reaction time factor (Factor C) has loadings of 0·73 for visual reaction time and 0·68 for auditory reaction times, but no other loading higher than 0·36. If either test had been excluded, Factor C would probably not have been isolated; and the communality for the remaining tests would have been much smaller.

The communality also varies according as the sub-tests are scored separately or as a single composite score. In Thurstone's study of ' primary mental abilities ' [27] tests of addition, subtraction, multiplication, and division were scored independently; but other investigators (e.g. the United States Army Air Force, [9], p. 80) have incorporated these into a single arithmetical test. On *a priori* grounds the composite might be expected to possess a larger common variance than any constituent sub-test, since it is longer and has greater reliability, But in practice it generally has a smaller communality. With one numerical score instead of four, variance previously regarded as common thus becomes classified as specific.

According to the mode of selection, therefore, it seems possible for the communality to range between zero and the reliability. Factorial theory can be restated in various ways to allow for this. First, we might declare that no two tests be included in the same analysis if too similar. Two reaction tests might be included, for example, when they relate to two different sensory modalities; but two flicker fusion tests would be excluded because both are visual. This seems hardly acceptable. The difficulty of deciding what is too similar is great; and investigators will seldom agree. Thurstone seems to have changed his mind about the similarity of the two flicker fusion tests after his administration of the tests. In any case, since one aim of factor analysis is to classify like tests together, it seems undesirable to prejudge the issue by classifying them before the analysis.

Secondly, the value attained for a representative finite selection of tests might be regarded as an approximation to the value which would be obtained if the analysis had been made for the whole domain. Guttman [11], [14] has investigated the consequences of this approach. If the number of common factors stays constant, then, as the number of tests grows larger, the communality will remain unchanged, but the s.m.c. will increase. The two will coincide when the number becomes infinite. Further, the s.m.c. will converge to the

communality even if the number of common factors increases, provided that the ratio of factors to tests decreases to zero as the number of tests increases to infinity. The practical difficulty is that it seems unlikely that the number of common factors will stay constant as further tests are added, or even that the ratio will in fact decrease. The ' upper bound ' for the rank of the correlation matrix, derived by Sheppard ([6], pp. 200) and Ledermann [22] from available degrees of freedom, often appears to give a reasonable estimate of the number of common factors required for an exact fit. Here, however, the ratio of common factors to tests increases to one as the number of tests increases to infinity. And even this ' upper bound ' may at times underestimate the number of factors. Guttman [17] provides one example, and the well-known Heywood case ([26], pp. 51–53) another. In a domain of tests, all variance hitherto regarded as specific might be expected to be converted to common, so that in the limit the communality would become equal to the parallel-form reliability.

Thirdly, we might make a virtue of necessity by stating our objective as that of designing analyses so as to convert as much specific variance to common variance as we can, thus narrowing the gap between communality and reliability so far as possible. To do so might be held to demonstrate our control of the testing situation and our power to express factorial results more generally. Or, fourthly, we might accept the relativity of common and specific variance on the ground that our concern is with the ' here and now '; in that case it would no longer be meaningful to refer to a test's communality as such; different communalities would be allowed in relation to different batteries. These last two alternatives, which accept a communality as a variable measure, probably represent the empirical situation accurately enough, and might prove acceptable operationally; but they destroy the notion of the reliable test variance as divisible once and for all into common and specific portions.

(iii) *Failure to Distinguish between Common and Significant Factors.* While the usual form of the theory allows for error introduced by unreliability, none is made for error arising from sampling of subjects. Thurstone never makes any clear distinction between a common factor, in the algebraic sense, and a significant factor, in the statistical sense. At times he implies that all error relates to self-correlations, and that the problem is solved once the reduced diagonal entries are known. But a common factor is not necessarily a significant factor. Suppose population values for all correlations are zero: there will still be some non-zero correlations in any sample, so that there will be at least one common factor. This will represent only sampling error and will not be significant. Nor is a significant factor necessarily a common factor in the Thurstonian sense. A principal axes analysis with unities in the diagonal also makes no distinction between common and specific variance: there are as many factors as tests; and, since later factors tend to have one or two loadings larger than the rest, they hardly conform to Thurstone's definition. The common variance is then so confounded with specific and error variance that such factors might be better labelled ' near-specifics '; yet they may still be significant: i.e. we may expect

similar near-specific factors on analysing another sample of the same size from the same population. Hotelling [18] and Bartlett [4] have devised significance tests for this situation.

When the problem is formulated in a purely algebraic sense such as that of reducing the correlation matrix to minimal rank, later factors may measure no more than sampling errors; yet this error will be included in the communalities and regarded as common factor variance. Our concern is therefore not with common factors *per se*, but only with those common factors that can be established as significant. As Burt has argued, " instead of aiming at a *maximum* number of common factors, a safer policy is to aim at a *minimum*, retaining merely those that are fully significant—*one* general factor only, if that will suffice to account for the figures observed, adding others only when the evidence compels us " ([8], p. 112).

Thurstone seeks to restate the problem in statistical terms by writing: " A matrix of minimal rank is to be obtained which deviates only slightly from the given side correlations " ([28], p. 283). But several difficulties arise from his formulation. First, it is doubtful whether the algebra of rank is intended as a guide to the number of factors, or whether his remark in the preceding paragraph that " the smallest number of independent factors that will account for the intercorrelations of n tests is the minimal rank of the correlation matrix " is meant to relate to the practical question. Second, Thurstone in his book, nowhere indicates the size of discrepancy which is small enough to be ignored.

(iv) *Dependence on the Sample of Persons and the Probability Level.* The problem of determining the number of significant common factors has been considered by Lawley [21] and Rao [25]. Increase in sample size usually results in reduced standard errors and in the attribution of significance (as the accepted probability level) to variations which would otherwise have been taken to represent merely sampling effects. It might be expected, therefore, that the number of factors would increase (*a*) if a more lenient level were adopted, or (*b*) if the sample of persons were increased. This in turn would lead us to identify as common some of the variance hitherto accepted as unique.

Is there any evidence that this situation fails to hold for common factors? Possibly the Thurstonian hypothesis of a limited number of common factors might be restated to apply to a population of persons, but (owing to random errors) not to a sample. As sample size increased, exact reduction of rank would be expected with fewer factors, instead of with more, because factors which resulted only from sampling error would grow smaller, and would disappear if correlations could be obtained for the population. On this view, increasing the size of sample would be offset by smaller residuals, and the number of significant factors would remain unchanged.

Two comments may be made. First, present evidence, though inconclusive, seems opposed to this view. Current neurological and genetic knowledge suggests greater complexity in the determination of the behaviour of the organism than can be accounted for by any faculty-type theory of a limited number of primary factors. With larger samples of persons and wider selections of tests

factorial results have become more and more complex. The theory of a restricted number of common factors might be tested empirically by finding whether the size of later factors diminishes as the size of the sample increases (implying a better fit to the common-factor hypothesis as sampling error diminishes). Secondly, this restatement requires that sample size be large enough to identify the full set of factors, but does not indicate how large a sample is needed. As we increase our sample, the number of common factors and the estimates of the communalities would change until the unknown number of common factors was reached.

As long as no concrete evidence is available in its support, we should perhaps view any hypothesis of a limited number of common factors with scepticism. It seems safer to regard the number of significant common factors as depending on sample size, with its corollary that the communalities will vary from one sample to another. If so, the assumption of a fixed division of reliable variance into common and specific portions again becomes untenable. As pointed out earlier, this relativity could be accepted by restating the objective as that of trying to narrow the gap between communality and reliability; in that case a large specific variance would be taken as indicating poor experimental design. A bigger sample might therefore be required to ensure (*a*) better determination of common variance, (*b*) more precise estimation of loadings, (*c*) isolation of small common factors otherwise unrecognised.

The most acute problem arises, however, because increase in the number of factors not only provides us with the additional information represented therein, but also modifies the structure of the remainder when factors are rotated. Suppose for the moment the objective to be finding diagonal values that yield residuals with the smallest sum of squares for an agreed number of factors. First, any increase in the number of factors may lead to (*a*) radical changes in diagonal values (they sometimes decrease when the number of factors increases), and (*b*) considerable changes in the factor loadings, with alterations in signs as well as the size, as Burt has pointed out ([8], p. 114). This source of restructuring is peculiar to the communalities formulation because their values are not known; the problem could be overcome by using some fixed value, e.g., s.m.c.'s, reliabilities, or unities, in the diagonal.

Secondly, when factors are rotated, increasing their number causes the larger factors to split into smaller: this happens even with physical measurements, e.g. when the quartimax procedure [24] is applied to the Pearson matrix analysed by Burt [8]. Unities in the diagonal may have increased this tendency; but Guilford ([10], p. 286) has also noticed it. No doubt it is responsible for some of the disagreement in the literature dealing with certain factors. The problem needs empirical investigation; but three observations may here be made.

1. To begin with, the problem is not confined to communalities. When the number of factors is made to depend on a test of significance, it becomes in part a function of size of sample, no matter what the diagonal entries may be. Indeed, changes in structure are likely to be less for communalities than for

unities or reliabilities (though greater than with s.m.c.'s), because reducing the diagonal values reduces the smaller factors proportionately.

2. Before including factors in a rotation, we might require them to be not only significant but also large enough to be psychologically important. Psychologists should certainly not rotate more factors than are significant; but it might be advantageous to operate with fewer. The map-maker does not try to put in all the known detail; and similarly the factorist might find loss of detailed information counterbalanced by increased clarity. Reduction might be achieved in one of two ways :

(*a*) We may introduce a rule for excluding factors which, though significant, are not of psychological importance. Kaiser [19] has proposed the exclusion of all factors for which the sums of the squares of the loadings are less than 1·00. It seems a doubtful gain to accept in the system a factor which contributes less information than any single test which it replaces. Tryon ([31], p. 17) suggests that the practical importance of a factor might be better judged by its highest loading than by its overall contribution; in that case we might reject every factor whose loadings were all under (say) 0·15.

(*b*) We may adopt that particular set of rotational solutions which achieves greatest clarity. This would require a definition of clarity and a way of measuring it : e.g. it might be argued that the clearest structure was achieved when the variance of the squared loadings in the rotated solution was greatest, the quartimax principle being used to decide the number of factors as well as the preferred rotation. Increasing the number of significant factors would then make no difference beyond the point where greatest clarity was attained.

3. The problem is greatly reduced if a hierarchical structure is adopted. Possibly because its basis lies in faculty psychology, simple structure theory has tended to regard each factor as coordinate in status with every other. Burt [7], followed by Vernon [32], considers a hierarchical organization to be far more effective than a coordinate one for classifying tests. This can be illustrated from the analysis of ability. If only one factor is extracted, a general factor of intelligence will probably be obtained. If two factors are extracted, verbal tests will probably be separated from spatial and mechanical. If further factors are extracted, the verbal and the mechanical factors will each subdivide. Thurstone's recent advocacy of second-order factors ([20], ch. 18) seems to represent a similar step towards hierarchical factors; so does Guilford's schematization of factors for thinking and memory [10].

A hierarchical system is akin to a genealogical chart: large factors are progressively subdivided into smaller. When the sample is small, few factors are significant, and only the higher levels can be established; with a larger sample, the chart can be filled in with greater detail. The rotation of greatest clarity is then represented by a horizontal segment across the chart.

III. Lack of Agreement in the Definition of Communality

One reason for all these difficulties is the lack of an accepted definition of communality. Most factorists agree about excluding specific and error variance

and invoking as few common factors as possible; but they differ as to how these aims are to be realized. Some, like Spearman, Thurstone, and Guttman, start with a hypothesis about the formal structure of the tests, and define the communalities as the values ensuring the best fit to the postulated model. Others, e.g. Tryon, and Guttman in some of his contributions, define the communality in terms of relations between a particular test and a domain. Others again, like Lawley and Rao, define the communality in statistical instead of algebraic terms[2].

Each of these alternatives deserves a brief examination. Until there is some agreement on the definition of communality, we can scarcely hope for any agreement about the values to be accepted.

The Spearman Model. Spearman assumed that, once specific and error variances had been removed, a single factor would suffice to describe the correlations. He himself scrupulously refrained from inserting any figures in the diagonal of his correlation tables. But, if his tables are to be factorized in the usual way, then the adoption of his theory would entail choosing for the communalities those values which most nearly reduce the matrix to rank one. To demonstrate the validity of the theory, it is necessary to test it with a wide variety of correlation matrices. This test, however, clearly revealed the need for rejecting the Spearman hypothesis: the majority of the correlation matrices actually obtained could not be reduced to rank one. Evidently the ' two factor theory ' was far too simple to accord with the facts.

[2]Pearson's method of analysis into ' principal axes ' assumes that the diagonal values will each be unity [8]. The first to substitute reduced values seems to have been Burt. In his earliest papers (*Brit. J. Psychol.*, III, 1909, pp. 94f. etc.) the tables of observed correlations have reliability coefficients (within brackets) inserted in the diagonal cells ; but to reconstruct the theoretical correlations to get the observed correlations he uses the formula $r_{ij} = r_{ig} r_{jg}$: so that $r_{ii} = r_{ig}^2$. The saturations themselves (r_{ig}, r_{jg}, etc.) are based on the sums of each column. In later papers he explains more fully the method of summation adopted to calculate the saturations. Where the reliability coefficients include a significant contribution from the specific factor (as in fact is most commonly the case), " a slight but obvious complication " (he says) " is necessary in applying formula iii " (the summation equation for the saturation coefficient), i.e., as the context indicates, it is necessary to reduce the reliability coefficient to a value agreeing more closely with the general trend [5, p. 53, footnote].

Thomson, quoting this and the corresponding passages in *Factors of the Mind* (pp. 448, 462) thinks it is still " not quite clear how he (Burt) filled in the blank diagonal cells " (26, p. 25). Burt, however, maintains that the use of a fixed rule (as with Thurstone's suggestion of using the highest correlation in the column) tends to yield misleading results. Since his own tables were always arranged as nearly as possible in hierarchical order, a process of ' smoothing ' seemed the quickest and most accurate way to obtain a provisional value ; and the value so obtained was adjusted (when necessary) by successive approximation. However, when the hypothesis to be tested was the adequacy of a single factor only, it was possible to modify the summation formula so as to omit the values for the ' self-correlations ' altogether. This proved desirable in discussions with Spearman, since he held that the method of successive approximation was " arguing in a vicious circle ". Unfortunately when more than one factor was assumed, no such device was available. Sheppard, however, pointed out that the single factor hypothesis was equivalent to postulating a matrix of rank one, and that, " when more than a single factor was present, the appropriate value would be the sum of the squares of *all* the saturation coefficients " ; and this was adopted as a convenient check in most of Burt's later papers (Burt [8] and personal communication).

The Thurstone Model. At present Thurstone's hypothesis is probably the most widely accepted. Reverting to faculty psychology, he postulated a limited number of ' primary mental abilities '. As we have seen, his theory was stated algebraically as that of attaining minimal rank for the correlation matrix[3]. However, the definition of communality in terms of rank has proved difficult arithmetically. For one thing, no easy method has been developed for finding minimal rank: Albert's procedure [1], [2] is too complex for regular use. Moreover both the number of factors and the values of the communalities have to be determined; and this double determination is all the harder because the two estimations are interdependent. Furthermore, if minimal rank be required in the literal sense of a solution that reduces all residuals to zero, then it would seem that the number of common factors can rarely be reduced below Lederman's ' upper bound ', and sometimes not even as low as that. Hence in practice far more factors have to be accepted than Thurstone apparently anticipated when formulating his theory. For example, when s.m.c.'s are inserted in the leading diagonal, the well-known Holzinger and Harman correlation matrix, based on 24 ability tests, has 13 positive latent roots: this means that at least 13 factors are needed to account for the correlations ([15], p. 153).

What is even more important, the actual factorial findings seem to be making Thurstone's hypothesis as unacceptable as Spearman's. Thurstone never stated explicitly how many factors in all might be expected. His adoption of the phrase ' primary mental abilities' in his analysis of cognitive tests [27] suggests that he expected a rather limited number; and his repeated insistence on the importance of parsimony supports this interpretation. Yet the number of factors empirically identified continues to increase. Guilford [10] lists no fewer than 40 in his review of work on thinking and memory alone, and expects to discover more when appropriate tests are constructed.

The Guttman Models. Guttman ([13], ch. 6) notes the inability of Thurstone's theory to fit the data, and has proposed new formal models such as the simplex, the circumplex, and the radex. These employ different relations among factors loadings from those ordinarily adopted (viz. accounting for correlations by the products of loadings). He accordingly defines communalities as the values most nearly maintaining the postulated structure. Up to the present, however, no allowance seems to have been made for errors resulting from sampling. As with Spearman and Thurstone, therefore, we can only accept Guttman's definition if his postulated structures are shown to be applicable to a wide variety of empirical matrices.

[3]Thurstone's account [29, pp. 282-283] is not altogether clear. Thus in a footnote on p. 283, he expresses his objective as finding the values which maintain the rank of the off-diagonal elements in the correlation matrix. This may have been to avoid certain criticisms from Lederman [22], who had demonstrated that anomalous correlation matrices might occur in which minimal rank was not associated with a minimal sum of communalities. It is difficult to give precise meaning to Thurstone's remarks about the rank of the off-diagonal entries unless we assume a minimal rank requirement. For similar cirticisms of this passage the reader may refer to this *Journal*, IV, p. 144, and VII, p. 108.

The Tryon Model. Tryon [20] formulates the problem quite differently. He defines the communality in terms of his domain-sampling theory, as the squared correlation between the test and a domain of tests whose correlations with the remaining $p-1$ tests are proportional. With this definition, the communality of a test can be shown to equal the correlation between it and a hypothetical test having identical correlations with the remaining $p-1$ tests in the selection. Correlations are then estimated between the tests and their hypothesized counterparts by psychometric formulae of the Spearman–Brown type. He thereby obtains a set of quadratic equations which when solved give exact values for the communalities ([30], p. 246)[4].

The procedure has been restated by Kaiser [20] in determinantal form; and the resultant iterative technique is found to converge for artificially constructed matrices. It is not yet known whether with empirical matrices convergence is slow or fails to occur. Mathematical properties of the method have also been examined by Guttman [15]. The conception of a test as drawn from a hypothetical domain of similar tests is attractive, since psychometric theory has usually paid more attention to the sampling of persons than to the selecting of tests. The power of the method is shown by the fact that it leads to explicit equations, and that their solution requires no knowledge or estimate of the number of factors.

On the other hand, to postulate a domain of tests with proportional profiles is only a necessary, not a sufficient, condition. There are further requirements unstated. For instance, parallel forms of a test, as encountered in reliability theory, would have proportional profiles, but would not meet the communality requirements, because drawn from the same behaviour domains rather than from different. The distinction between communality and reliability, so Tryon would hold, is that the domain for the reliability must be defined in advance: items drawn randomly from this predesignated domain would provide the requisite parallel tests; but the domains for the communality are operational fictions defined after the tests have been selected for analysis.

As we have seen, alternative sets of communalities are possible; and it is by no means clear why Tryon's procedure should lead to the minimal rank set, though it certainly seems to do so, at least with artificially constructed matrices. Possibly this follows from the particular form which the linear dependence of each common-factor portion on all the others assumes. Because of this assumption of rank reduction, the practical consequences of the Tryon–Kaiser method seem to be mostly similar to Thurstone's, and thus subject to the same criticisms. First, no allowance is made for errors arising from the sampling of persons.

[4]Using the same concepts, Tryon has developed several different methods for determining the communalities: see [30] and [31]. Kaiser derives his determinantal equation in terms of Thurstone's common-plus-unique partitioning rather than of Tryon's correlation with a hypothetical test having an identical correlational profile. He argues that the rank approach of Thurstone and the domain of Tryon lead to the same arithmetic and the same practical consequences. In his view, although a formulation in terms of rank leads to the simpler algebra, the logic based on the 'domain' concept yields a fuller insight from the standpoint of psychology.

An algebraic solution is designated in which all residuals are reduced exactly to zero. Hence a large number of factors will generally be needed in any small sample—more perhaps than might reasonably be expected to be meaningful. Secondly, Tryon explicitly admits that communalities will change as extra tests are added. ' Proportional tests ' are merely fictional devices, introduced in order to achieve a more compelling conceptualization; it is not asserted that such tests could be developed empirically. But definition of the communality in such terms means that the profile has to be extended to cover further tests added to the selection.

For Tryon, therefore, the communality is always restricted to the context of the tests in a particular battery. Within that context he considers that, except for sampling variations, the communality will have only one value. Since he acknowledges that diagonal entries should be based on the ' here and now ', i.e. on the particular selection of tests chosen by the investigator, one wonders whether a case might not be made for using the actual s.m.c. which measures variance held in common with the remaining $p-1$ tests, instead of a fictional s.m.c. This will be discussed later.

The Statistical Model. The final position to be considered is that in which the problem is stated in statistical rather than algebraic terms. This is represented by Burt's procedure of correcting for uniqueness ([8], p. 113), Lawley's maximum likelihood method [21], and Rao's canonical method [25]. The communalities are then the values which enable the correlations (weighted in accordance with the communalities) to be fitted as well as possible for the significant factors, i.e. they are those which give the smallest sum of squares of weighted residuals.

This includes some Thurstonian features, but rejects others. The number of factors depends on significance, not on rank; but rank reappears in the need to find diagonal values providing the best fit between correlations and factor, i.e. which achieve rank reduction, not to the minimum, but to that implied by the indicated number of significant factors; and these diagonal values are used in applying significance tests. Many disadvantages of Thurstone's algebraic approach are overcome by such methods. The requirement that common factors should also be significant factors is obviously in line with modern statistical theory. Furthermore, the method is free from some of the theoretical difficulties of Thurstone's position. The distinction between common and unique variance is no longer an attribute of a test *per se*, a position hardly tenable in view of variability of the communality : instead it is made relative to the particular set of tests and persons, and its justification resides simply in the fact that the loadings provide the best fit to the correlations. We are no longer committed either to the faculty theory of a limited number of factors or to the postulate that their number either stays constant when tests increase, or increases at a slower rate.

Two difficulties remain. The first is logical. While dependence of the communality upon the test selection may be psychologically acceptable, it is hard to give psychological meaning to the variations in the communality that arise

from changes in the sample of persons. The approach seems to imply that in a very large sample practically all common factors will be significant, so that the values for the communalities would then approach those that would be secured by minimizing rank. The maximum likelihood communality is therefore a biased statistic. Samples nearly always give too low a value, because communalities are estimated, and factors are tested for significance, in a single calculation. The two problems might be separated by first determining communalities algebraically, e.g. by taking the minimal rank values or values conforming to Tryon's definition. These might be regarded as estimates for the selection of tests for a population of persons. There would remain an error of estimate depending on sample size; but there would be no inevitable bias on the low side. Having calculated the factors derived from these communalities, we could then test the hypothesis that any particular factor is not different from zero. The estimational procedure would thus be separated from the test of significance.

In Rao's procedure there is a second weakness. For the computer the iterations for determining the number of significant factors and the best diagonal entries are very laborious, and in general are practicable only with an electronic computer. This may not prove a permanent difficulty. Calculations might perhaps be accelerated by better initial estimation and by extrapolation to speed convergence. Moreover, faster electronic computers may soon be available.

IV. Communality, Reliability, and Squared Multiple Correlation

The central argument of this paper has been that the distinction of common variance has not been shown to be stable. This section will consider the consequences if communalities should prove relative to the selection of tests and persons.

The Communality as a Domain Measure of Common Variance. Thurstone and Tryon agree in regarding the communality as a measure of variance common to a test and its domain, but define the domain differently. Thurstone postulates a domain of tests with the same factorial composition as those already selected. He does not state this explicitly; but, as Guttman has shown, it follows from his hypothesis. Since no satisfactory evidence has so far been provided to prove that the number of factors stays constant, insistence on an unchanging boundary between common and specific variance seems unwarranted. Can Thurstone's position be modified to meet this difficulty? The distinction between 'common' and 'specific' might be made one of degree by conceding that further common factors can be obtained by adding tests, while maintaining that, if the selection of tests was representative at the outset, the further factors will be small and of little psychological relevance. But such a restatement seems unconvincing. First, even small factors might lead to an appreciable change in a rotated solution because of additional dimensions. Secondly, since the theory specifically implies that the communality is a measure of the common variance of a domain, even the small factors should not be excluded in estimating it:

otherwise calculated values would be merely an uneasy compromise between the value for a domain and that for the particular selection.

Tryon, on the other hand, postulates a different domain for each test. Thus in a different way he requires, like Thurstone, that no new common factors be introduced in determining the common variance in the domain. The appearance of new common factors he meets by redefining the domains. Once this is conceded, it would seem to follow that, by adding a sufficiently large and representative set of tests, all specific variance will be converted to common. In the limit, therefore, the communality would equal the reliability: by reliability we here mean the correlation between a test and a parallel form, since a parallel form could be added as a further test. A thoroughgoing adoption of domain-sampling therefore seems logically to require abandoning the common–specific distinction altogether. In the domain, test variance will be divisible into two parts only, reliability and error, so that reliabilities should be inserted in the leading diagonal.

The Communality as the Value Giving the Best Fit to the Observations. An alternative case can be made for communalities on the ground that they provide the smallest residuals. Thurstone apparently hoped that communalities would give the best correlational fit, as well as measure common variance in a domain. This would be providential. But the coalescence could only be achieved by the unproven assumption of a limited number of common factors in the domain.

The correlations merely offer a summary description of the relations between the test measurements. Hence if the aim is to fit the observed data as closely as possible, it would seem better to seek to reproduce the test measurements: i.e. if there is to be a reduction from p tests to k factors, the factors should be those for which the factor measurements are most highly correlated with the test measurements. This is the approach of the maximum likelihood method. Factors so secured are those which provide the maximal canonical correlations with the tests; and Rao accordingly prefers to call the procedure the *canonical* method of factor analysis ([25], p. 94). The disadvantages of this formulation, if we may judge from certain results when using the maximum likelihood method, seem to be the possibility of obtaining diagonal values which do not conform to customary expectations: e.g. when the sample is so small that only one factor is significant, a canonical correlation between tests and factor of 1·00 can be secured by reproducing one test as the factor. More generally, if the problem is to maximize the relations between known tests and unknown factors which are functions of the tests themselves, something resembling spurious correlation may conceivably arise. This seems to hold for all communality formulations. Since the factor loading is the correlation between test and factor, the communality is equal to the correlation of any test with the full set of common factors. The test is therefore to be correlated with factors of which it forms part. Such a theory seems to leave us free to arrange the conditions of the right-hand side of the test-factor relation so that we can "pull ourselves up by our own bootstrap".

The Squared Multiple Correlation as a Measure of the Common Variance in the Test Selection. The possibility of using s.m.c's deserves greater consideration than it has so far received. Both in theory and in practice the communality has proved a somewhat intractable concept. If s.m.c's. were inserted in the leading diagonal, the distinction between common and specific variance would be based on the test selection instead of a hypothetical domain. It is well known that the s.m.c. is a lower bound for the communality. The s.m.c. measures variance common to the test and the remaining $p-1$ tests in the selection, whereas the communality measures that common to the test and the factors based upon the p tests. The additional ' common variance ' introduced by using communality instead of s.m.c. is therefore a measure of the extent to which the test itself is reproduced by the factors. Can this fairly be called common variance ? Common variance initially was considered to be variance common to two or more tests, but the communality is now seen to include a certain amount of variance that is common only to the test and a transformation of itself. It is scarcely psychologically meaningful to call this additional portion ' common variance '. The dilemma could have been avoided had it been shown that the increment represents extra common variance obtained by increasing the selection to the limit ; but this line of argument leads to the reliability instead of to the communality.

Various objections raised against communalities can be met, or be partially met, by the use of s.m.c's. The problem of alternative sets of diagonal values satisfying the theory then disappears. S.m.c's are probably less influenced by increasing the sample of persons, since the revised values, though more accurate, may either be higher or lower. Determining the diagonal entries becomes separate from determining the number of significant factors. One disadvantage is the lack of a significance test ; but this is a problem for the statistician rather than the psychologist.

The common–specific distinction is still relative to test selection, since s.m.c's increase when further tests are added. Hence knowing the common variance would provide no information about the test *per se*, transferable to other analyses, but would indicate the relation between that test and the other tests included in the analysis. Marked discrepancies between s.m.c.'s and parallel-form reliabilities might suggest defects in test selection. Computationally there is an attractive precision about the s.m.c.'s. On any high-speed machine they can be obtained much more easily than communalities.

Two related problems need investigation : (*a*) the size of the increments from s.m.c's to communalities ; (*b*) the extent to which factorial results are changed by using s.m.c's for communalities. In default of more direct evidence, it must suffice to consider the logical alternatives and their implications. At one extreme there is the possibility that increments remain constant from variable to variable : the two sets of principal axes loadings would then be proportional, the only difference being that the latent roots are greater with the latter than with the former by the value of the constant ; so that with

communalities the minor factors play a larger part in rotations. In this case therefore the use of s.m.c's would make little difference in practice ; but the s.m.c's would be simpler computationally.

The key issue is then represented by the other extreme, namely, where increments differ greatly from one variable to another. Here communalities have definite algebraic advantages, provided that they can be accurately determined : (*a*) the correlation matrix is then Gramian, i.e. there will be no negative latent roots with associated imaginary loadings ; (*b*) the diagonal values with which the analysis starts are reproduced by the factor loadings. But even here the psychological advantages seem to rest with the s.m.c's. Consider the artificial correlations in the table below. In the first case the communalities are approximately proportional to the s.m.c's ; in the second the factor loadings have been chosen so that there will be a marked discrepancy between the s.m.c. and the communality for the fifth variable. For this the correlations in the second case are only 0·045 larger than in the first ; yet we are asked to believe that

To Illustrate the Effects of Minor Changes in Factorial Structure on the Communalities and the Squared Multiple Correlations

Factor loadings		Correlations					Communalities	Squared multiple correlations
0·90	—	—	0·72	0·63	0·54	0·09	0·810	0·622
0·80	—	0·72	—	0·56	0·48	0·08	0·640	0·546
0·70	—	0·63	0·56	—	0·42	0·07	0·490	0·426
0·60	—	0·54	0·48	0·42	—	0·06	0·360	0·315
0·10	—	0·09	0·08	0·07	0·06	—	0·010	0·009
0·90	0·05	—	0·7225	0·6325	0·5425	0·1350	0·812	0·626
0·80	0·05	0·7225	—	0·5625	0·4825	0·1250	0·642	0·549
0·70	0·05	0·6325	0·5625	—	0·4225	0·1150	0·492	0·429
0·60	0·05	0·5425	0·4825	0·4225	—	0·1050	0·362	0·318
0·10	0·90	0·1350	0·1250	0·1150	0·1050	—	0·820	0·022

in the first case the common factor variance is only 1 per cent of the test's total variance, whereas in the second case it amounts to 82 per cent. If someone unfamiliar with factorial theory were confronted with these two cases, he would surely regard the s.m.c. as offering the better measure of the ' commonness ' of the fifth variable. No assertions that the communality is the value attained by the s.m.c. in a domain of tests would affect his verdict. Rather he would believe that Thurstone had allowed the algebra of the situations tc dominate over the psychology.

S.m.c.'s, of course, do not answer all the issues raised in this paper. Thus, if significant factors alone are rotated, their number may increase as the sample of persons increases ; and that may lead to a radical restructuring. The fact

that the factorial results may very possibly change with increase in the sample of persons, or in the selection of tests, should warn us that any hope that factors can be regarded as real causal entities is unlikely to be realized. Factors must be regarded as operational fictions. Our task is to fix the rules governing their extraction so that our practical problems, e.g. the classification of the tests, can be most skilfully and most expeditiously handled. And it certainly seems possible that s.m.c's may serve these purposes better than communalities.

V. Summary

This paper considers the distinction between specific and common factor variance and the possibility of separating them by using communalities. If such a distinction is to provide information about the tests, it seems essential to demonstrate the stability of the communality from one battery of tests to another and from one sample of persons to another.

There are three principal reasons for believing that the actual values found cannot be stable. First, the existence of alternative sets of communalities in any one analysis requires some further restriction to decide between them ; and any such restriction seems to depend more on human convenience than on any fundamental test characteristic. Secondly, standard theory seems to have made inadequate allowance for sampling errors when the same test battery is factored for different-sized samples of persons. To achieve generality, factors must be significant as well as common. The number of factors and the size of the associated communalities will therefore depend on the number of persons in the sample. Thirdly, difficulties arise when further tests are added to the battery. If no new common factors were introduced as a result, the communality for any test would remain constant ; but empirical findings do not support this expectation. Illustrations have been given which show (*a*) how the addition of appropriate tests will convert specific to common variance, thereby increasing the communality, and (*b*) how removal of tests will reduce common variance to specific.

Once the distinction between common and specific variance is recognized as operational, i.e. as relative to the particular testing situation, we are led to ask whether the communality is the most effective instrument for reducing the number of factors. If our concern is with common variance in a domain, a better case could be made for reliabilities, since all the specific variance can in theory be converted to common by adding enough tests. The principal claim of communalities rests on the fact that they give a better fit than any other diagonal values to the correlations actually observed. The psychological advantages, however, appear to rest with the squared multiple correlations, which measure the proportion of common variance in any particular test selection.

References

[1] Albert, A. A. (1944). The matrices of factor analysis. *Proc. Nat. Acad. Sci.*, XXX, 90–95.

[2] Albert, A. A. (1944). The minimum rank of a correlation matrix. *Proc. Nat. Acad. Sci.*, XXX, 144–146.

[3] Banks, C., and Burt, C. (1954). The reduced correlation matrix. *Brit. J. Stat. Psychol.*, VII, 107–118.

[4] Bartlett, M. S. (1950). Tests of significance in factor analysis. *Brit. J. Psychol. Stat. Sect.*, III, 77–85, IV, 1–2.

[5] Burt, C. (1917). *Distributional and Relations of Education Ability*, London : P. S. King.

[6] Burt, C. (1952). Bipolar factors as a cause of cyclic overlap. *Brit. J. Psychol. Stat. Sect.*, V, 197–202.

[7] Burt, C. (1949). The structure of the mind ; a review of the results of factor analysis. *Brit. J. Educ. Psychol.*, XIX, 100–111, 176–199.

[8] Burt, C. (1949). Alternative methods of factor analysis and their relations to Pearson's method of ' Principal Axes '. *Brit. J. Psychol. Stat. Sect.*, II, 98–121.

[9] Guilford, J. P., ed. (1947). *Printed Classification Tests.* Washington, D.C. : U.S. Govt. Printing Office. Army Air Forces Aviation Psychology Program Research Report, No. 5.

[10] Guilford, J. P. (1956). The structure of intellect. *Psychol. Bull.*, LIII, 267–293.

[11] Guttman, L. (1940). Multiple rectilinear prediction and the resolution into components. *Psychometrika*, V, 75–99.

[12] Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, XVIII, 277–296.

[13] Guttman, L. (1954). A new approach to factor analysis : the radex. In Paul F. Lazarsfeld, *Mathematical Thinking in the Social Sciences.* New York : Columbia University Press.

[14] Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, XIX, 149–161.

[15] Guttman, L. (1956). Best possible systematic estimates of communalities. *Psychometrika*, XXI, 273–285.

[16] Guttman, L. (1956). Improved bounds for communalities. *Research Report.*

[17] Guttman, L. (1956). To what extent can communalities reduce rank ? *Research Report.*

[18] Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.*, XXIV, 417–441, 498–520.

[19] Kaiser, H. F. (1956). The varimax method of factor analysis. Ph.D. Thesis, on file in the University of California Library.

[20] Kaiser, H. F. (1956). Solution for the communalities : a preliminary report. *Research Report.*

[21] Lawley, D. N. (1940). The estimation of factor loadings by the method of maximum likelihood. *Proc. Roy. Soc. Edin.*, LX, 64–82.

[22] Ledermann, W. (1937). On the rank of the reduced correlation matrix in multiple factor analysis. *Psychometrika*, II, 85–93.

[23] Ledermann, W. (1940). A problem concerning matrices with variable diagonal elements. *Proc. Roy. Soc. Edin.*, LX, 1–17.

[24] Neuhaus, J. O., and Wrigley, C. F. (1954). The quartimax method : an analytic approach to orthogonal simple structure. *Brit. J. Stat. Psychol.*, VII, 81–91.

[25] Rao, C. R. (1955). Estimation and tests of significance in factor analysis. *Psychometrika*, XX, 93–111.

[26] Thomson, G. H. (1951). *The Factorial Analysis of Human Ability.* London : University of London Press.

[27] THURSTONE, L. L. (1938). *Primary Mental Abilities*. Chicago : University of Chicago Press.
[28] THURSTONE, L. L. (1944). *A Factorial Study of Perception*. Chicago : University of Chicago Press.
[29] THURSTONE, L. L. (1947). *Multiple Factor Analysis*. Chicago : University of Chicago Press.
[30] TRYON, R. C. (1956). Communality of a variable : formulation by cluster analysis. *Research Report*.
[31] TRYON, R. C. (1956). Cumulative communality cluster (CCC) analysis. *Research Report*.
[32] VERNON, P. E. (1950). *The Structure of Human Abilities*. London : Methuen.

Vol. X Part II The British Journal of Statistical Psychology November 1957

THE RELATIVE INFLUENCE OF HEREDITY AND ENVIRONMENT ON ASSESSMENTS OF INTELLIGENCE

By Cyril Burt and Margaret Howard

The Application of the Parallelogram Law. In several previous publications we have briefly referred to a graphical method of representing and computing the effects of genetic and non-genetic factors which is based on the same principle as the familiar diagram used for illustrating the composition and resolution of forces in physics—the diagram popularly known as the ' parallelogram of forces '. The method we propose treats the variables concerned as measurable tendencies or ' vectors ' obeying the ordinary rule for the addition of vector quantities. It has proved specially helpful in attempts to analyse and interpret the factors involved in measurements of intelligence. Results obtained in this way were cited in a previous paper on multifactorial inheritance in this *Journal* (IX, pp. 126–7). Several correspondents have since asked for a more detailed description of the procedure. We therefore offer a short account of the working method, illustrated by data given in the article just quoted.

In explaining the principles on which the calculations are based, we should like to emphasize at the outset that we regard the values so reached as representing only approximate estimates for the hypothetical variables concerned. The approximation is, we believe, sufficiently exact for use with measurements of the kind we have in view, since their error or unreliability is in any case likely to be rather large ; but we do not claim that the method will furnish the most efficient estimates, such as might satisfy the ideals of the mathematical statistician. Indeed, Mr. Alan Stuart, who has been good enough to read and comment on an early draft of this note, states that in his opinion the type of problem here raised is one which still awaits an adequate theoretical investigation.

Hypothetical Analysis. In applying the procedure to the problem mentioned we may start by dividing the conditions which affect our assessments of intelligence into the following hypothetical factors.

I. *Genetic or Heritable Factors* (H). These include (i) the *Fixable* contributions resulting from ' the essential genotypes ' ; (ii) the *Non-fixable* contributions due chiefly to the effects of partial dominance ; and in addition to these it is usually desirable to allow for (iii) the effects of *Assortative Mating*. (Formal definitions of these terms are given in the article already quoted, p. 106.) For our present purpose it will not be necessary to deal with these subclasses separately.

II. *Non-genetic or Non-heritable Factors* (N). These include (i) *Random Environmental Effects*, i.e. the effects of environment so far as it operates in complete statistical independence of the genetic tendencies (R), (ii) *Systematic Environmental Effects*, i.e. the effects of environment so far as it operates in the same direction as the genetic tendencies (S), (iii) a contribution due to the

Interaction of heredity and environment (if any), i.e. to the joint effects of both when they are correlated (*J*), (iv) all other unidentified influences—errors of measurement, sampling errors, etc., which may be summed up under the general term *Unreliability* (*U*).

Data. The data available (collected from a sample of nearly 1,000 children) consist of (*a*) the crude marks obtained from group tests of the usual type (verbal and non-verbal), (*b*) the adjusted assessments obtained by submitting the initial marks for criticism or correction to teachers and then adjusting the marks where necessary by supplementary tests of an individual and non-verbal type. In addition we have assessments, obtained from somewhat smaller groups, for the apparent intelligence of the parents, of the nearer relatives, and of twins, as well as ratings for the economic and cultural conditions of a representative sample of families.

From these various measurements or assessments we have computed the following coefficients :

(i) Correlations between two successive tests or assessments : these enable us to assess the amount of ' unreliability ' ;

(ii) Correlations between identical twins who have been (*a*) brought up together and (*b*) separated from infancy and brought up apart : these enable us to estimate the effects of post-natal environment (intra-twin variance).

(iii) Correlations between offspring of first cousins : these by exclusion should allow us to estimate the possible effects of pre-natal environment ; in point of fact, however, the effects appear so small as to be virtually negligible.

(iv) Correlations between (*a*) parents and offspring, (*b*) siblings, (*c*) fathers and mothers : these enable us to assess the relative influence of the first three ' heritable ' conditions separately (if required) as compared with the influence of environment.

Preliminary Analysis of Variance. Using these correlations, and relying mainly on the formulae derived in our previous paper, we can partition the total variance for each set of assessments as follows (Table I : for the detailed working see *loc. cit.*, IX, pp. 124f).

Table I. Analysis of Variance

	Crude Test Marks	Adjusted Assessments
Heredity	77·06	87·56
Environment : random effects	5·91	5·77
Environment : systematic effects	10·60	1·43
Reduced Total	93·57	94·76
Unreliability	6·43	5·24
Total	100·00	100·00

Analysis by Vectors. We now propose to regard the main components as vector quantities (i.e. quantities defined solely by their magnitude and direction)

and to treat the observed assessments as a resultant obtained by combining the components in accordance with the ' parallelogram law ' for the addition of vectors, on the analogy of the resultant and component forces in mechanics.

In the problem we are examining here we are not concerned with the effects of unreliability. These form a random variable uncorrelated with the remaining tendencies, and would therefore have to be represented in a third dimension. Here we shall be concerned solely with an attempt to partition what we may call the ' reduced total ', i.e. the total left after the effects of unreliability have been deducted. Accordingly, on the basis of the first three percentages shown in Table I we may now re-analyse the data in the following way.

So far as direct causation is concerned the variations in a child's genetic constitution are completely independent of the variations in his environment. But, as a matter of empirical fact, it is always conceivable that in any selected group the variations of the two factors may be in some degree correlated, possibly owing to some indirect process of causation. Let us therefore begin by taking as our two rectangular axes, first the horizontal line OH, to represent, by its direction and length, the relative influence of heredity (H), and a second line perpendicular to it, namely, the vertical axis, OR, to represent the influence of environment (R) in so far as it is uncorrelated with heredity (see Fig. 1). Their lengths will accordingly be drawn proportional to the square roots of the corresponding variance, i.e. to the standard deviations.

We now require a supplementary component to represent environment, in so far as it operates in the same direction as heredity. This can be represented by the line RS parallel to OH. The joint effects of these two environmental influences will then be indicated by a line, OS. Evidently the observed test-marks, OT, will be found by completing the parallelogram, OSTH.

Problem. As analysed in Table I, the results obtained yield figures indicating (i) the length of the base, OH, (ii) the length of the diagonal, OT, and (iii) the height of the parallelogram, OR. The point T can therefore be determined by describing a circle with centre at O and radius OT cutting RT in T. Hence our main task will be to determine the lengths of the lines RS and OS (which are needed to measure the influence of the components S and J and can be calculated by Euclid I, 47) and the angles between the various vectors (which are needed to measure the various correlations).

Working Procedure. The suggested steps in the calculations are simple if a little lengthy. To begin with, we find the lengths of the lines OH, OR, and OT by taking the square roots of the variances : e.g. for the crude test measurements (col. 1 of Table I), OH = $\sqrt{(77{\cdot}06)} = 8{\cdot}7784$; OR = $\sqrt{(5{\cdot}91)} = 2{\cdot}4305$; OT = $\sqrt{(93{\cdot}57)} = 9{\cdot}6732$.

Since cos ROT = OR/OH = 0·2513, we can at once determine the angle ROT (75° 26′) and its sine (0·9679); and, since RT = OT sine ROT, we can now calculate RS = RT − ST = RT − OH = 9·3627 − 8·7784 = 0·5843. The remaining steps are obvious. We compute in turn : tan ROS = 0·2403, angle

ROS = 13° 29′, cos ROS (r_{rs}) = 0·9724, SO = 2·5002, angle SOT = 61° 57′, cos SOT (r_{st}) × 0·4703, angle SOH = 76° 31′, cos SOH (r_{sh}) = 0·2332, angle TOH = 14° 34′, cos TOH (r_{th}) = 0·9679. The square of OS gives the variance of the joint environmental effects, viz. $2{\cdot}5002^2 = 6{\cdot}2514$; and the last three cosines specify the three correlations which we chiefly desire to know—namely, those between (i) the systematic effects of the environment and the tests, 0·4703, (ii) the systematic effects of the environment and heredity, 0·2332, and (iii) heredity and the tests, 0·9679.

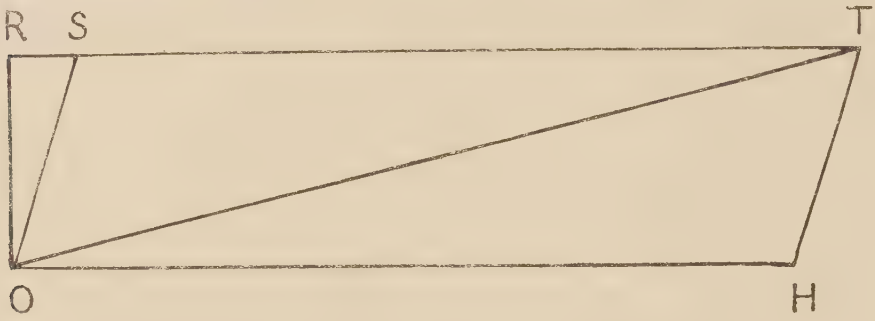

Fig. 1. Heredity (OH), Environment (ST), and their Resultant (OT).

Alternative Analysis of Variance. We can now re-analyse the variance of the reduced total, σ_t^2, in terms of the chief contributory factors just specified. Writing σ_h^2 for the variance contributed by heredity, σ_j^2 for the variance of the joint environmental effects, and r_{hj} for the correlation between the two, we have, by the familiar formula,

$$\sigma_t^2 = \sigma_h^2 + \sigma_j^2 + 2r_{hj}\,\sigma_h\,\sigma_j;$$

that is, for the Crude Test Marks,

$$= (8{\cdot}7784)^2 + (2{\cdot}5002)^2 + 2 \times 0{\cdot}2332 \times 8{\cdot}7784 \times 2{\cdot}5002$$
$$= 77{\cdot}06 + 6{\cdot}25 + 10{\cdot}26 = 93{\cdot}57.$$

In the same way we obtain for the Adjusted Assessments,

$$(9{\cdot}3574)^2 + (2{\cdot}4031)^2 + 2 \times 0{\cdot}0317 \times 9{\cdot}3574 \times 2{\cdot}4021$$
$$= 87{\cdot}56 + 5{\cdot}78 + 1{\cdot}42 = 94{\cdot}76.$$

We may tabulate these results as shown in Table II.

Table II. Alternative Analysis of Variance

	Crude Test Marks	Adjusted Assessments
Heredity	77·06	87·56
Environment	6·25	5·78
Effects of correlation between heredity and environment	10·26	1·42
Total	93·57	94·76

This procedure enables us to subdivide the total variance in such a way as to show not only the proportion contributed (i) by genetic influences and (ii) by environmental influences (both ' systematic ' influences, i.e. those operating in the same direction as heredity, and ' random ', i.e. those operating in statistical independence), but also (iii) the contribution due to the joint effects arising from the fact that the two main components may be more or less correlated. We observe that, for the crude test measurements, this third contribution is fairly large—over 10 per cent, whereas for the adjusted assessments it is almost negligible, amounting to between 1 and 2 per cent only.

Empirical Verification. From the initial figures for the component variances, it will be noted, we are able by this method to deduce an estimate for the correlation between environmental influences, on the one hand, and the assessments for intelligence, on the other. Taking the crude test marks, we obtain a correlation of 0·470; and on substituting the adjusted assessments we obtain 0·277. It should be possible to check these two theoretical estimates by comparing them with the empirical coefficients. All that we require are objective assessments for the relevant environmental conditions. It was impossible to procure the necessary figures in the case of every individual child in the entire group. But, as noted in our previous paper, we were able to secure trustworthy ratings for the ' socio-economic status ' of a large number of households, which together form a fairly representative sample. These include ratings for (*a*) the material (i.e. financial and hygienic) conditions of each family, and (*b*) the cultural (i.e. educational and motivational) background.

(*a*) For the material environment the correlations with the assessments of intelligence are (i) with the crude test marks 0·288 and (ii) with the adjusted assessments 0·226. (*b*) For the cultural environment the correlations are (i) with the crude test-results, 0·453, and (ii) with the adjusted assessments, 0·315. These later coefficients agree quite well with the theoretical values we have deduced; and they thus provide an added confirmation for the validity of the analysis.

Causal Influences and Path-Coefficients. Finally, we may note that the square roots of the two variances, represented by the lines OH and OS, may be written $\beta_{\text{th.e}}$ and $\beta_{\text{te.h}}$, and can be regarded as the partial regression coefficients of the actual assessments (i.e. in the diagram, of the test results) on heredity and environment respectively. Sewall Wright terms such quantities ' path-coefficients ': these two, it should be observed, are ' one-way ' path-coefficients. The correlation between heredity and environment, r_{he}, he would also regard as a ' path coefficient ': but it is essentially a ' two-way ' coefficient. With this interpretation we can think of heredity as influencing (or as capable of influencing) the test-results along two paths simultaneously : first, by the direct path represented by the regression coefficient $\beta_{\text{th.e}}$, and secondly by an indirect path, leading first from heredity to environment (as indicated by the presence of a significant correlation between them), and thence by the path leading from environment to the test-results as indicated by the other regression coefficient, $\beta_{\text{te.h}}$. An

analogous interpretation may be adopted, at least as a theoretical possibility, in regard to environment; environment may influence the test-results both directly and indirectly, the indirect influence passing first through the same two-way path as before, but in the opposite direction.

However, when we attempt to examine causal influences in this concrete way, it seems scarcely satisfactory to stop at this point. There are two obvious questions which still remain undecided. Both are questions which statistical analysis *by itself* is utterly incompetent to solve.

In the first place, as we have noted, environment may in some degree operate, or tend to operate, in the same direction as heredity. Suppose we could show that this effect occurred merely among those families where the parents, in virtue of their own inheritable intelligence, had been able to provide a better cultural environment for their children; many would be tempted to argue that, instead of crediting this effect to environment, we ought to regard it as yet another instance in which heredity is the primary cause. Whether such a statement is reasonable would depend on how the initial question itself was phrased; and no doubt, in the sweeping form we have just used, the conjecture would almost certainly be misleading: sometimes parents of no great intelligence do succeed, either by making financial sacrifices or (in times past) by the aid of inherited wealth, in providing exceptional intellectual opportunities for their children; these opportunities would almost inevitably increase the *apparent* intelligence of their children (though perhaps not so much as is commonly believed). This is not the place to discuss such issues in and for themselves. We mention them to show that they are of a type which can only be decided by specific investigations, and that in such investigations statistical analysis will play merely an auxiliary part.

Secondly, as the suggestions of several correspondents indicate, many feel that it would be desirable to split what we have called "the effects of the correlation between heredity and environment", and assign appropriate portions definitely to one or the other of the two correlated variables. In dealing with cases like the present, Heilman, for example, has suggested that the joint contribution should be divided in proportion to their direct contributions [1]. This seems rather arbitrary. We ourselves believe that, without further data obtained expressly to settle such an issue, it is impossible to carry the analysis further than we have done; and the requisite data could only be procured by a line of inquiry quite different from that with which we are here concerned. In any case the results recorded in the last of our tables (Table II) seem to give a reasonably adequate answer to the limited questions with which we began; and these complex supplementary issues must be deferred to a later occasion.

[1] Heilman, J. D., 'The Relative Influence upon Educational Achievement of Some Hereditary and Environmental Factors', *XXVIIth Yearbook Nat. Soc. Study Educ.* (1928), pp. 35–65; cf. *id.*, 'Factors determining Achievement and Grade Location', *J. Genet. Psychol.*, XXXVI, 1929, pp. 435–456.

Vol. X Part II The British Journal of Statistical Psychology November 1957

THE THEORY AND USE OF THE SUCCESS-RATIO

By David A. Walker
Deputy Director of Education, Fife

Abstract. The success-function defined in this paper has been found useful in follow-up studies of school children where entrance marks have been normally distributed. It provides a rapid method of estimating the product-moment correlation between entrance mark and success when the criterion for success is of the pass-or-fail type.

I. Introduction

No system of selecting pupils for courses of secondary education is, or can be, completely successful while children preserve their right to develop in their own ways and at their own speeds. But in predicting success in the secondary school different batteries of tests show different degrees of efficiency; and much of the research in this field has consisted of follow-up studies undertaken to compare the validities of tests and other measures used at the selection stage. The concept of success-ratio is one that does not appear to have been used to its full extent; and the present paper is an attempt to develop its theory and application.

The assumptions that must be satisfied if a full use is to be made of it are: (i) that the entrance marks of the whole group under consideration are normally distributed; (ii) that success (or its lack) is a normally distributed variable, those pupils above a certain level in that attribute being successful, while those below that level are unsuccessful; (iii) that the correlation between entrance mark and success is linear.

The applicability of the first assumption to any set of data can readily be tested, while the second and third assumptions can be checked by the methods described in this paper. It will be observed that the question of correction for attenuation is not raised. The method thus avoids the difficulties both of suitable assumptions and of additional calculation experienced by previous experimenters (e.g. W. G. Emmett [1], W. McClelland [2]).

II. Theory of the Success-Ratio

If the assumptions just stated are satisfied, it is possible to calculate the expected value of the success-ratio for any given entrance mark. Let x represent the entrance mark, measured in units of standard deviation from zero mean. Let y represent the success–failure variable, similarly measured. Let $y = k$ be the line dividing success from failure, values of y greater than k representing success. The correlation surface is

$$Z = \frac{1}{2\pi\sqrt{(1-r^2)}} \exp\left\{ - \frac{1}{2(1-r^2)} (x^2 - 2rxy + y^2) \right\},$$

where r is the correlation between the variables x and y. For a given value of x, the frequency of success is therefore

$$\int_k^\infty Z\,dy = \frac{1}{2\pi\sqrt{(1-r^2)}}\int_k^\infty \exp\left\{-\frac{1}{2(1-r^2)}(x^2-2rxy+y^2)\right\}dy,$$

which can be shown to equal

$$\frac{\exp(-\frac{1}{2}x^2)}{2\pi\sqrt{(1-r^2)}}\int_k^\infty \exp\left\{-\frac{1}{2(1-r^2)}(y-rx)^2\right\}dy.$$

Now $\exp(-\frac{1}{2}x^2)/\sqrt{(2\pi)}$ is the total number of candidates with mark x in the group. Hence

$$\frac{\text{Number successful at level } x}{\text{Total number at level } x} = \frac{1}{\sqrt{(2\pi)}\sqrt{(1-r^2)}}\int_k^\infty \exp\left\{-\frac{1}{2(1-r^2)}(y-rx)^2\right\}dy$$

This fraction may be termed the 'success-ratio' at x.

Let $$\frac{y-rx}{\sqrt{(1-r^2)}} = -u, \qquad dy = -\sqrt{(1-r^2)}\,du.$$

Limits of integration are

$$-u = \frac{k-rx}{\sqrt{(1-r^2)}} \text{ and } \infty, \qquad \text{i.e. } u = \frac{rx-k}{\sqrt{(1-r^2)}} \text{ and } -\infty.$$

Hence the success-ratio at level x

$$= \frac{1}{\sqrt{(2\pi)}\sqrt{(1-r^2)}}\int_{\frac{rx-k}{\sqrt{(1-r^2)}}}^{-\infty}\{-\sqrt{(1-r^2)}\}\exp(-\tfrac{1}{2}u^2)\,du$$

$$= \frac{1}{\sqrt{(2\pi)}}\int_{-\infty}^{\frac{rx-k}{\sqrt{(1-r^2)}}}\exp(-\tfrac{1}{2}u^2)\,du. \quad (1)$$

It is therefore the integral of a normally distributed variable with zero mean and unit variance between the limits $-\infty$ and $(rx-k)/\sqrt{(1-r^2)}$. In other words, it is a function of r, x and k only. Consider this as a function of x. Put

$$u = \frac{ru'-k}{\sqrt{(1-r^2)}}, \qquad du = \frac{r}{\sqrt{(1-r^2)}}\,du'.$$

Limits of u' are now $-\infty$ and x, and success-ratio at level x

$$= \frac{r}{\sqrt{(2\pi)}\sqrt{(1-r^2)}}\int_{-\infty}^{x}\exp\left\{-\frac{(ru'-k)^2}{2(1-r^2)}\right\}du', \quad (2)$$

which is the integral of a normally distributed variable u' with mean at $(ru'-k)/\sqrt{(1-r^2)} = 0$, i.e. $u' = k/r$, and standard deviation $\sigma' = \sqrt{(1-r^2)}/r$, both measured in σ units from mean of ellipse.

This function with mean at k/r and standard deviation equal to $\sqrt{(1-r^2)}/r$ we shall term the 'success-function'. The theorem established is that the success-ratio at any extrance mark, x, standardized on the whole group is the integral of the success-function from $-\infty$ to x.

III. Applications

(*a*) *The rapid estimation of sucess-ratios.* This application is most readily illustrated by an example. Suppose that the entrance marks are I.Q.s centred on 100 with standard deviation 15, and that the correlation between I.Q. and success is 0·8; suppose also that the line dividing success from failure is at $k = 1$ S.D., i.e. 16 per cent succeed.

Then the success-function is the probability function with mean at $k/r = 1/0{\cdot}8 = 1{\cdot}25$ corresponding to an I.Q. of $100 + 1{\cdot}25 \times 15$ or 119, and standard deviation of $\sqrt{(1-r^2)}/r = 0{\cdot}75$ corresponding to an I.Q. range of $0{\cdot}75 \times 15$, i.e. 11.

From this distribution the theoretical success-ratio at any precise I.Q. level can be estimated. For example, pupils with I.Q. levels 119 have a success ratio of 0·50; those with I.Q. 130, i.e. 11 points above this mean, have a success-ratio of 0·84; those with I.Q. 108 have a success-ratio of 0·16, and so on. These values may be read from tables of the normal distribution, or from a straight line drawn on probability paper.

Table I. Success-Ratios of 1949 Group at University Entrance Level

Entrance Mark	Number in Group	Number gaining University Entrance	Success-Ratio
100 and over	8	5	0·625
95–99	29	12	0·414
90–94	105	27	0·257
85–89	225	32	0·142
80–84	377	10	0·026
75–79	491	3	0·006
70–74	514	2	0·004
Below 70	2581	0	—
Total	4330	91	—

Mean Mark 65·3. Standard Deviation 14·3.

(*b*) *Estimation of correlation from success-ratios.* The usual situation is that the success-ratios are known, but not the correlation between entrance mark and success; and the experimenter wishes to estimate the correlation from the data. The discussion will be clearer if an example is given from actual practice. In 1949, 4,330 pupils in a Scottish county were transferred to courses of secondary education, their courses having been partly determined by the entrance mark in a battery of tests and other measures; five years later, 91 had obtained a Scottish Leaving Certificate of a sufficiently high standard to admit them to a Scottish University. The performance of the whole group is shown in Table I. To save space the lower part of the table has been telescoped; the distribution of entrance marks was in fact roughly, though not completely, normal.

Method A. From the theory given above it is obvious that an estimate of the correlation coefficient r can be obtained from each success-ratio separately,

provided that k is known. For example, the second group which has a mean entrance mark of 97 has a success-ratio of 0·414. The upper limit of u (eqn. (1)) is $-0{\cdot}217$; hence

$$(rx-k)/\sqrt{(1-r^2)} = -0{\cdot}217. \qquad (3)$$

Now $x = (97-65{\cdot}3)/14{\cdot}3$ in standardized units; and k is the line cutting off 91/4330 from the tail of the normal curve, i.e. $k = 2{\cdot}034$. (Inspection of Table I suggests that no suitable candidate has been excluded from the successful group by the operation of the pass mark. This point will be considered in greater detail in a later section.) The solution of eqn. (3) (taking the positive value of the root since it represents a standard deviation) is $r = 0{\cdot}87$. Estimates of r could be obtained in the same way from the other entrance levels, and these could be combined. The most accurate way of doing this is shown under Method C below.

Method B. The most rapid method of estimating the correlation coefficient from a set of success-ratios is by the use of the relation between r and the standard deviation of the success function. If the success-ratios are plotted on probability paper against the mid-points of the entrance-mark intervals, the points should lie on a straight line if the assumptions underlying the method are satisfied. From this straight line, the entrance mark for which the success-ratio is 0·50 and that for which the success-ratio is 0·16 are read off. These give the mean of the success-function and the point one σ unit below the mean. The difference between these measures is therefore the standard deviation of the success function. When it is standardized by division by the standard deviation of the total group of entrance marks, it equals $\sqrt{(1-r^2)}/r$; and r is rapidly found.

In practice the points are not absolutely collinear; and a line of best fit may be drawn by eye. When this is done with the data of Table I, it is found that a success-ratio of 0·50 occurs at entrance mark 98·0 and one of 0·16 at entrance mark 89·5. Hence

$$\sqrt{(1-r^2)}/r = (98{\cdot}0-89{\cdot}5)/14{\cdot}3, \quad \text{i.e. } r = 0{\cdot}86.$$

Alternatively, we may convert the success-ratios into the corresponding normal deviates; and the graph may then be drawn on ordinary graph paper. If the assumptions are satisfied, the graph is once more a straight line, and the standard deviation of the success-function corresponds to unit change in the normal deviate.

Method C. For a more exact method of estimating the correlation coefficient from a set of success-ratios we must fit the best line to the series of points, giving each set of observations the appropriate weight. This is ensured by the method devised by Bliss (cf. Fisher and Yates, *Statistical Tables* for *Biological, Agricultural and Medical Research*, Introduction to Tables IX and XI). Here we are concerned with the success-ratios of pupils with different entrance marks. Bliss's method not only enables us to calculate the slope of the line relating the normal deviate to entrance mark, and so to calculate the correlation underlying the success-ratios, but also supplies a value for chi-squared indicating whether the assumptions of normality and of linear correlation are in agreement with the data.

Details of the calculations are given in the Appendix (pp. 46f below). When the method was applied to the data of Table I, the correlation coefficient obtained was 0·85. The value of chi-squared was 5·8, giving a probability of at least 0·5 that the deviations could be attributed to sampling errors.

IV. The Effect of the Pass-Mark

If entrance to the course is limited to those with entrance marks $x > h$, then the successful candidates will be those in the correlation ellipse for whom $> h$, $y > k$. The theory outlined in section II of this paper applies only to those for whom $x > h$, and modifications must be made in the methods of estimating r. Method A is no longer applicable, as k is not known. Methods B and C can still be used for all groups above the admission mark.

Let us apply Method C to the data of Table II which shows the number of pupils from the 1949 group who obtained a certificate of any type after six years of secondary education. The admission mark to the course was 75, but some pupils with lower marks were admitted on appeal and others transferred to the Leaving Certificate course after having done well in other courses. A dotted line is drawn across the table to separate these candidates from the others.

Table II. Success-Ratios of 1949 Group in Minimum Certificate after 6 Years

Entrance Mark	Total Number in Group	Number Successful	Success-Ratio
100 and over	8	8	1·000
95–99	29	19	0·655
90–94	105	56	0·533
85–89	225	102	0·453
80–84	377	91	0·241
75–79	491	56	0·114
.....	.....	.....	.....
70–74	514	10	0·020
65–69	543	2	0·004
60–64	568	1	0·002
Below 60	1470	0	0·000

When Method C is applied to the figures above the dotted line, the value of r obtained is 0·79 and the value of chi-squared is about 5·5, giving a probability of 0·25 that the deviations are within the bounds of sampling errors. Let us now suppose that no candidate below the pass mark of 75 had been admitted to the course. The numbers successful in the class intervals 70–74 and below would then have been zero in each case. The application of Method C would furnish a correlation coefficient of 0·86; but the value of chi-squared would become 27·7, and the consequent probability about 0·001. The deviations would be too great to be attributed to sampling errors and the assumptions, e.g. linearity of correlation, would be no longer tenable. No importance could be attched to the correlation so obtained.

It is interesting, though rather irrelevant to the main purpose of this article, to note that when the actual numbers of successes are used, the value of r obtained from the whole table is 0·84, and the value of chi-squared is 12·7; which corresponds to a probability of 0·20. The flexibility of the selection scheme has therefore restored the position. The correlation obtained from the whole table is greater than that obtained from the part above the line, but it is not possible to state whether the difference is significant, as the standard errors of correlation coefficients obtained by these methods have not been worked out.

V. Summary

1. From a set of success-ratios there can be derived a success-function which can be used to predict other success-ratios and also to estimate the correlation between entrance mark and success.

2. Methods of estimating the correlation between entrance mark and success are given and illustrated by examples.

3. The advantages of the use of the success-function in estimating correlations are :

(i) the assumptions are few, simple, and easily tested;

(ii) the intricacies of correcting correlation coefficients for attenuation are avoided.

APPENDIX. Details of Calculations for Method C

For ease of calculation the mid-points of the class intervals of the entrance marks may be denoted by the variable x, chosen so that $x = 0$ at entrance mark 87. For each success-ratio we find the appropriate probit, i.e. the normal deviate increased by 5 to avoid negative values. These probits may be obtained from a table of the normal distribution; a convenient table is provided in Fisher and Yates' Table IX. They may be described as the experimental probits, and are shown in Column 5 of Table III. We next plot the experimental probits against the entrance mark x and draw the best fitting line by eye. From this line we obtain provisional probits for use in the next part of the calculation. The provisional probits chosen should be simple values, occurring in Fisher and Yates' Table XI. We now calculate from the data given in Table XI the working probit y. This equals the minimum working probit plus the success-ratio times the range. For example, the working probit for $x = 3$ is $3{\cdot}7186 + 0{\cdot}625 \times 2{\cdot}5573$.

Table XI also gives the weighting coefficient to be used with each provisional probit. The actual weight w is obtained by multiplying the weighting coefficient by the total number of pupils at that level.

We now calculate $S(w)$, $S(wx)$, $S(wy)$, $S(wxy)$, $S(wx^2)$, and $S(wy^2)$, and from them obtain

$$\bar{x} = S(wx)/S(w) = -0{\cdot}3408;$$
$$\bar{y} = S(wy)/S(w) = 3{\cdot}593;$$
$$A = Sw(x-\bar{x})^2 = S(wx^2) - \bar{x}S(wx) = 537{\cdot}549;$$
$$B = Swy(x-\bar{x}) = S(wxy) - \bar{x}S(wy) = 305{\cdot}708;$$
$$C = Sw(y-\bar{y})^2 = S(wy^2) - \bar{y}S(wy) = 179{\cdot}684.$$

TABLE III. DATA FOR DETERMINING b AND χ^2

Entrance mark	x	Total Number	Number Successful	Experimental Probit	Provisional Probit	Working Probit (y)	Weight (w)
100–104	3	8	5	5·3	5·2	5·3169	5·016
95–99	2	29	12	4·8	4·7	4·7831	17·864
90–94	1	105	27	4·3	4·2	4·3563	52·815
85–89	0	225	32	3·9	3·7	3·9650	75·600
80–84	−1	377	10	3·1	3·2	3·0809	67·860
75–79	−2	491	3	2·5	2·7	2·5371	37·120
70–74	−3	514	2	2·3	2·2	2·3698	12·644
65–69	−4	543	0	—	1·7	1·4194	3·334
60–64	−5	568	0	—	1·2	0·9522	0·670

The slope of the line relating probit to entrance mark is $b = B/A = 0{\cdot}5687$. Unit change in probit therefore corresponds to a change of $1/0{\cdot}5687$ in entrance mark. This must be standardized in terms of the class interval of 5 that has been used and of the standard deviation of the whole group of entrance marks, i.e. 14·3. The corresponding change of entrance marks is therefore $5/(0{\cdot}5687 \times 14{\cdot}3)$ or 0·6148 which thus represents the standard deviation of the success function. This, according to the theory, equals $\sqrt{(1-r^2)}/r$, from which $r = 0{\cdot}85$.

Also $\chi^2 = C - bB = 5\ 83$. The appropriate number of degrees of freedom is 7, since two constants have been fitted, and the corresponding value of P is between 0·5 and 0·6.

REFERENCES

[1] EMMETT, W. G. (1945). *An Inquiry into the Prediction of Secondary School Success.* London : University of London Press.

[2] MCCLELLAND, W. (1942). *Selection for Secondary Education.* London : University of London Press.

JOURNAL COMMITTEE

As instructed by the Council of the British Psychological Society, the Journal Committee, at its meetings in 1956, carefully re-examined the questions of title, typography, and price; and now wishes to thank the numerous contributors and readers of the journal who helped by communicating their views. In its final recommendations the Committee strongly supported the new arrangements for printing, and urged that there should be no change either in the title or in the price of the journal. It suggested that the arguments of those critics who found the articles too technical might be partly met if contributors could be persuaded to submit papers on a wider range of topics, e.g. experimental, elementary or expository articles, and notes or abstracts from theses.

Council deferred a full consideration of the Committee's recommendations until its meeting on July 6, 1957, i.e. until after the first issue of the current year had been published. The decisions then reached were reported by the Editor to the Committee in his review for the first half year of 1957 (to July 31), which the Committee recommended should be published.

The report stated that the changes in typography, which the Editor had proposed in 1954 (11-pt Imprint with 1-pt leading, and abolition of small type) had now been sanctioned. Largely as a result, the previous deficit had been converted into a profit; and accordingly, after some discussion, Council had agreed that the price should at present not be raised.

Council, however, feared that subscribers might now feel that they were getting an inadequate return for their subscriptions, since increase in type-size and spacing, following on the temporary reduction to two issues a year, would mean that less material could be presented in an issue of 64 pages. The Editor explained that it was now proposed to return to the plan of issuing three parts a year, and that a benefactor had guaranteed a fund to cover any loss that could not be met out of the profits. Part III would include an Index for the first ten volumes. The alternative would be to increase the number of pages in each part. The former proposal was accepted by Council. But the Editor would still welcome the views of readers.

Since the Publications Committee had recommended that editors should serve for a limited period of years only, Council had nominated Dr. Whitfield as his successor.

It was agreed that a formal expression of thanks should be recorded to the American benefactor for her aid, to contributors and readers for their continued help and support, and to all those (including the new printers and publishers) who had enabled the journal to reach a position where no serious deficit need any longer be feared.

Correspondents are asked to note that the Officers of the Society are no longer the Publishers of the Journal. Communications regarding material for publication should be sent to the Editor; questions relating to business matters should be addressed to the Managing Editor.

CHARLOTTE BANKS,
Secretary.

Vol. X Part II The British Journal of Statistical Psychology November 1957

CRITICAL NOTICE

Secondary School Selection : A British Psychological Society Inquiry. Edited by P. E. VERNON. London : Methuen & Co., Ltd., 1957. Pp. 216. 15*s*.

This survey will be widely welcomed for its summary of recent literature on the important issues raised : and the conclusions put forward may no doubt be taken as representing the general views of British psychologists on one of the most heated controversies of the day[1]. The book deserves the special attention of the statistical psychologist, not only because it shows how serviceable his techniques have proved, but still more because it sets forth a number of specific problems on which his help will be essential.

The report is the joint production of a Working Party which consisted of about a dozen members of the British Psychological Society. With few exceptions each contributed one or more chapters ; and, as was almost inevitable, the treatment varies appreciably in clarity and outlook. Thanks no doubt to the care taken by the Editor and Secretary (Professor Vernon and Miss Doris Lee), the statistical issues are clearly brought out and the statistical evidence given full weight.

The Preface warns the reader that " the memorandum cannot represent the views of every member of the Society ; and no attempt is made to conceal that on some points differences of opinion or emphasis have arisen ". In certain cases a non-committal compromise appears to have been accepted. In others the contradictions are left unresolved ; and it is frankly admitted that " divergences of outlook remain ". As several critics have already pointed out, it is perhaps to be regretted that the Working Party included none of the many experienced school teachers who are members of the Society nor any educational official or administrator whose memory could reach back to the practical difficulties which hampered the early reformers and which are so largely responsible for the shape the problem takes today. One consequence is that the issues involved are discussed as if they were primarily psychological, whereas it is the administrative requirements that must determine any workable solution[2].

[1] In view of the intense interest which the problem has recently aroused and of the volume of correspondence received by the editors since the Society's inquiry was published, the Journal Committee suggested that the book should be reviewed at some length. The task has been entrusted to two writers who are fully acquainted with modern statistical methods and have also had first-hand experience of the practical issues either as teacher or as administrator. The Editor regrets that space does not permit the publication of the letters in this issue : but the reviewers have been asked to take note of the points so raised. The full report recently published on the same subject by the National Foundation for Educational Research was received too late for review at the same time ; but it will be discussed in the following number. The reviewers would welcome further correspondence, which will be published or summarized at the same time.

[2] Miss Lee, who has been good enough to read this review in manuscript, points out that the terms of reference received by the Working Party referred primarily to '*psychological aspects* of secondary school selection' (*Editor*).

The Inquiry starts with a discussion of the ' Background of Selection ', and then passes to a ' History and Survey of Present Procedures '. These two chapters (as one of our correspondents observes) " have apparently been drafted by writers who are more familiar, it would seem, with the recent findings of American psychologists than with the earlier efforts of English investigators ". No reference is made to the work of such leading educationists as Ballard, Kimmins, Sir Robert Blair, Sir Percy Nunn, and Sir Michael Sadler, whose endeavours to place ' secondary school selection ' on a scientific basis were far more influential than the few who are mentioned by name. The bibliography of nearly 150 numbers provides a useful conspectus of recent publications. But only ten authors are cited who made contributions to the problem before 1930 ; and of these, six are American.

The writer of the first chapter tells us that it was " the Butler Education Act that gave us the blue print for the future of our schools which won unusually unanimous approbation "; and he apparently supposes that the proposals then made represented " an attempt to catch up with the changes in the structure of society following the war ". In point of fact, there was hardly anything in the Act which had not been suggested twenty years before ; and much of the present difficulty is due to the fact that even among progressive educationists the attitude towards those suggestions has been far from unanimous. It would be truer to say that the ' blue print ' was originally provided by the most celebrated of all the Reports of the Consultative Committee of the Board of Education —the report on *The Education of the Adolescent* (1926). Yet, strangely enough, the list of references at the end of the book omits all mention of it.

The inspiration behind that report came largely from Sir Percy Nunn, Principal of the Institute of Education, and by far the most eminent educationist of his day. It was he who so effectively urged the substitution of statistical studies and statistical data for the impressionistic arguments that had previously prevailed; it was his papers and memoranda which had argued so persuasively for the " clean cut across our educational system at the age of 11 + as in accordance with the results of recent psychological research "; and it was he and his colleagues at the Institute who pressed for the use of psychological tests as an aid to selection and secured the abolition of the ' Rule 3(b) ' of the Board's *Regulations for Secondary Schools*, which had for so long hindered their adoption. For many years he had urged[1] that the term ' secondary ' should be extended to cover *all* types of ' post-primary education ', and that a third type of secondary school, in addition to the ' technical ' and ' grammar ', should be introduced, namely, " schools with a realistic or practical trend to be known as ' Modern Schools ' " (a name suggested as equivalent to the German term *Realschulen*). And it was his suggestion that, in order to be sure of catching in the net the ablest members of each group, the proportion actually transferred to the grammar schools should be raised to between 15 and 20 per cent—the actual figure varying according to the population of the area. At the same time both Nunn and the Consultative

[1] T. P. Nunn, *Education: its Data and First Principles* (1920), pp. 237f.

Committee envisaged many other types of school or class according to the requirements and traditions of each district.

We cannot therefore help feeling that the Working Party has represented the problem as much newer and far simpler than it really is. Moreover, in all its reports the Consultative Committee treated its various proposals as involving a process of *allocating* pupils to suitable schools rather than of *selecting* them for favoured types of school. Unfortunately, as the title of their ' Inquiry ' implies, the Working Party seem rather to have retained the popular notion of the procedure, namely, that it is a kind of competitive selection for grammar schools. " At least half the child population," they say, " is competing for grammar schools which will only accommodate one-fifth." But that, plausible though it may seem to many, suggests a decidedly misleading picture of the fundamental problem. The real need at the moment is not so much more grammar schools for those who at present cannot get in, but more schools of a technical, commercial, or other specialized types. And, although at the end of the historical survey there is a short ' Note on Selection for Technical and Commercial Schools ' reminding the reader that " this topic must not be omitted altogether ", this aspect of the matter, and the numerous psychological questions which it involves, are in point of fact never really faced.

Having outlined the general problem and described in somewhat vague terms ' the Diversity ' and ' the Adaptability of English Education ', the writer of the opening chapter goes on to ask why after all parents are so anxious to see their boys and girls accepted by a grammar school. His answer is that, when their child fails to win acceptance, parents from the middle classes " feel that they are losing face ", and as a result often prefer a relatively inefficient private school to the " stigma of going to a modern school ". Teachers who have enjoyed a wider first-hand knowledge of the reasons that actuate most parents apparently think this explanation a little unfair. Thus several of our correspondents are of the opinion that (to quote one writer) " snobbery as a motive is far more frequent among the manual classes than among parents of the professional or middle classes: how often does one hear a father say: ' I only let him go there to show that he's as good as any other kid; I never meant him to stay ' ". As Mr. Skeet has said, owing perhaps to some lurking sense of inferiority, the working class parent " does not so much want a *suitable* education for his child, but only that his child shall go to as good a school as their neighbour's, i.e., to the one that is most esteemed "[1]. On the other hand, middle class parents (if we may trust a survey recently carried out by Mr. R. T. Graham) tend as a rule to adopt a far less egoistic attitude: with them the commonest objection to the modern school is, " not that it entails a social disgrace, but rather a fear lest the child may pick up rude manners, a low vocabulary, a slovenly accent, slangy and ungrammatical speech, and above all disgusting words and ideas ". By going to a grammar school with a long-standing tradition, the boy will (so the parent hopes) not only escape such risks, but " acquire something of the cultural,

[1] D. V. Skeet, *The Child of Eleven* (University of London Press, 1957).

moral, and social ideals, the spirit of loyalty and sense of responsibility and public service that are part of the very atmosphere of the English public schools and old-established grammar schools ".

In their history of the psychological study of the problem the writers are evidently more at home. Yet even here they seem to understimate the work done in this direction by the more progressive administrative authorities themselves—notably the inquiries and reports undertaken by the London County Council and by the Board of Education itself. It is to these bodies and to certain outstanding members of their staffs—Young, the Secretary of the Consultative Committee, Daniel, the Chief Examiner to the London County Council, Kimmins, its Chief Inspector, and Nunn, the Principal of its Training College (later known as the Institute of Education)—far more than to the academic writers whose names are better known, that the real initiative was chiefly due.

As the Preface points out, the contribution of the psychologist as such starts with " the pioneering work of Galton, who was interested in discovering innate abilities of a high order as well as in assessing cases of subnormality ". But the systematic use of intelligence tests as part of the educational routine began with the London County Council—the first education authority to appoint a psychologist as an Officer on its educational staff. The duties of the psychologist were not merely to apply psychological methods, but to investigate their practical value and, where necessary, develop and recommend new procedures.

For the assessment of subnormal children, the most obvious procedure was to use ' individual ' tests, such as those devised by Galton and Binet. For the scholarship examinations, intended to select brighter children for central schools and grammar schools, the application of individual tests (the only type available in the early days) would have been far too expensive; and accordingly a scheme of written tests for the majority, checked by teachers' ratings, with oral examinations for borderline cases, was put forward for trial. The idea of employing ' group ' tests started with the experiments on ' higher mental processes ' carried out by Burt in 1909; and he was apparently responsible for most of the stock types still used in examinations for secondary schools and described in the Working Party's report—Opposites, Analogies, Classification, Reasoning, Number Series, short Completion tests, and the like.[1] In view of recent complaints of coaching and ' test-sophistication ', it is surprising that so few fresh types have been invented during the past forty years. Here, we feel, the Working Party might well have suggested new types of problem, or at least emphasized the need for devising tests of a less antiquated kind.

In these earlier years most academic psychologists looked on such proposals with marked disfavour. Spearman urged the substitution of " simple laboratory tests, as befits the rigour of scientific work " (e.g. tests of sensory discrimination individually applied); Thomson questioned the whole notion of a general factor of intelligence. And the prolonged controversies over the very existence

[1] Cf. Experimental Tests of Higher Mental Processes and their Relation to Intelligence, *J. Exp. Pedag.*, I, 1911, pp. 93–112.

of ' general ability ' and of ' special aptitudes ' served only to augment the natural scepticism of officials and committees[1]. Nevertheless, the experiments carried out by the London County Council aroused widespread interest among administrators; and, as soon as the first world war was over, the group tests standardized by Burt for London were tried by education authorities in Blackpool, Leicester, Northumberland, the West Riding of Yorkshire, and elsewhere, as well as for entrance examinations at Cheltenham Grammar School and at Rugby. All who used them pronounced favourably on the results. But the first authorities to introduce a separate intelligence test as part of their regular procedure were those whose areas included a mixture of both rural and urban schools: as one of the reports explained, " in country districts many pupils may be quite as well endowed in natural ability as those in the towns; but, for various reasons, notably the low standards often prevailing in the village schools, they are apt to reach a comparatively poor level in ordinary examinations of a scholastic type ".

In 1920 the Consultative Committee took up the problem at the request of the Board. As regards the theoretical issue the Committee unhesitatingly accepted the distinction between ' inborn ability ' and ' acquired attainments ', and the double assumption of an ' innate general ability ' and ' innate special aptitudes maturing rather late '; and concluded that " a number of children qualified to profit by secondary education are at present excluded because, for one reason or another, their attainments are not equal to their abilities . . . . . Examinations for entrance to secondary schools should be designed primarily to discover ability rather than to test attainments ". Accordingly " tests of ' intelligence ' " were recommended for adoption, in addition to other methods, on the ground that they provided " a species of written or oral examination designed to test ability rather than attainments ".

It is evident, however, that the Committee were not entirely satisfied with the use of the word ' intelligence ' to describe the tests used for this purpose. The reasons are perhaps most clearly set forth in the memorandum by Burt referred to in their report. A distinction is there drawn between " the various purposes for which so called intelligence tests are used: in examining the borderline defective it is desirable, before certifying the child, to ascertain, so far as practicable, whether his backwardness is the irremediable consequence of an innate lack of general ability or merely the incidental result of remediable causes; but in examining a pupil for admission to a secondary school, it is less important to assess his innate ability as such, since that is but one of many factors determining his capacity to profit by the education there given. For instance, the verbal and scholastic bias of many tests, which should be wholly eliminated in tests for the defective, may actually prove an advantage in selecting pupils for a comparatively bookish and academic type of education . . . . It is, therefore, unfortunate that the phrase ' intelligence test ' is used to describe two different types of test having two different functions. The teacher in the elementary school, when receiving new pupils at the age of 7 or 8, rightly seeks a method of

[1] See Board's report on *Psychological Tests of Educable Capacity* (1924), especially appendix IX, which was largely concerned with these underlying issues. However, by the time the report was printed, both Thomson and Spearman had begun to modify their earlier views.

testing innate ability, since at that stage what he wants to know are the inborn potentialities of each child; but that is no longer the primary purpose of scholarship examinations at the age of eleven or later." Here " the real criterion is, not, how accurately does this procedure enable us to assess inborn potentiality, but how accurately does it enable us to predict the child's capacity to benefit by the type of education proposed ". The Working Party state that the situation now obtaining in Britain " clearly calls for some estimate of the child's potentialities at the end of the primary stage ". We should prefer to see this done at the beginning. And for the test used to estimate the child's suitability for the grammar or modern school we should prefer some such name as ' general classification test '.

The equivocality of the term ' intelligence ' is responsible for much confusion. As the Working Party go on to remind us, " intelligence suggests something inherited; thus the parent is much more emotionally involved in the child's I.Q. than in his performance at attainment tests or ordinary examinations ". Accordingly one of their chief aims has apparently been to allay these parental fears, which are the source of so much ill-informed opposition. They seem to have thought that the best way to do so would be to minimize the importance of hereditary differences and the possibility of determining them. But after all it is the existence of these innate differences that causes the problem and makes it of such national importance; and surely one can deny the need to test a thing for practical purposes without denying its existence on theoretical grounds.

The chapter which deals with the actual use of so called ' intelligence tests ' is much more lucidly written than the rest. The author frankly accepts the results of factor analysis, but rejects the extreme and antagonistic views of both Spearman and Thurstone. " Though factor analysis does not enable us to pin down the nature of intelligence," he says, " it has nevertheless developed to a stage where it can effectively analyse and classify the component abilities . . . . The factor g (for *general* ability) is," we are told, " an empirical fact: as Burt pointed out, it is the Highest Common Factor that can be extracted from analysing any set of test scores "; but at the same time " he provided evidence for numerous group factors or subtypes of ability." In spite of Spearman's arguments therefore we must recognize both.

The writer then goes on to inquire " how far do intelligence tests measure inborn ability ? " This, however, as we have already implied, is really a side-issue; and in any case the answer must surely be that all depends on which test we have in mind and on the way it is applied. On the general problem several extensive investigations have recently been published in the Society's own journals (cf., this *Journal*, IX, p. 126 and refs.) to which the Working Party does not refer. It would appear that with a good group test of the ordinary type about 77 per cent of the variance is attributable to genetic factors; and with supplementary checks even this figure can be raised. For evidence on this issue, however, the Inquiry relies mainly on earlier American research. In the end the views of " left-wing writers " who hold that " intelligence test results largely or

mainly reflect social class differences " are firmly repudiated; and Lawrence's studies of children transferred to orphanages are quoted as " proving beyond doubt that such social class differences [as the tests reveal] are in some measure genuinely innate ".

So far, therefore, the conclusions seem to agree both with those of the Consultative Committee and of the ordinary teacher[1]. But, as we read on, we encounter those unreconciled " divergences in outlook " that we are warned to expect. The Preface, as we have seen, urges the need for estimating the child's innate potentialities, and describes how education authorities are now using " tests designed to give an indication of innate ability as distinct from scholastic performance ". Yet, on a later page we are assured that such tests " do *not* measure innate ability ". Again in another passage, drafted presumably by another hand, we are told that " by intelligence we mean the more general qualities of comprehending, reasoning, and judging, which have been *picked up* without much specific instruction "; and it is therefore argued that, if I.Q. tests give a fair indication of capacity for advanced education ", that is " not because the capacity is inborn, but because it has been more or less stabilized in the previous eleven years ". It is in our view perfectly true to say that what we want to test in such examinations is ' stabilized capacity for further education '; but to accept this view does not mean that we must first deny genetic differences or the possibility of assessing them.

The mention of ' I.Q. tests ' leads to a discussion of the time-honoured question of the constancy of the I.Q. " From about 6 to 11 ", it is said, " the I.Q. from a good individual test is unlikely to alter by more than about 7 points on the average "; (that should reassure teachers who wish to determine at an early age which children are innately handicapped and which have the potentialities of good scholars). Hence " one can definitely conclude that the I.Q. is sufficiently stable to make possible useful predictions in the majority of cases over the period of primary or secondary schooling ". But once again, having thus confirmed the general view, the writers then proceed to make such large concessions to the rival doctrine that the ordinary reader is bound to feel bewildered. In certain cases, it is said, the I.Q. " may change 30 or 40 points; . . . many children with I.Q.s in the 50 to 70 range may show themselves capable later of coping with education in the ordinary school ". Let us apply this to children tested at 10: it means that a pupil who then had a mental age of 8 might at 12 have a mental age of over 14, and that " many " who had mental ages between 5 and 7 could later be satisfactorily educated in the ordinary school. Surely in such cases the more reasonable inference would be that the I.Q.s were obtained either by an incompetent tester or with unsuitable tests.

[1] A few years ago Professor Vernon submitted a questionnaire to a number of teachers, asking (among other things) how they understood the term ' intelligence ': the definition ' inborn all round mental efficiency ' (or its equivalent) was accepted by 84 per cent and rejected by only 9 per cent, the rest declining to commit themselves: (*The Bearings of Recent Advances in Psychology on Educational Problems*, 1955).

But all this is really irrelevant. Whether or not group tests of the usual type are of practical help in the selection examination is an issue to be decided, not by inquiring how far " the I.Q., properly measured, is stable or not ", but by ascertaining how far the tests actually do predict the children's subsequent performances. And to the statistical reader the portion of greatest interest will be the chapter dealing with " validity "—a problem which for some reason is discussed before the description of the tests and other procedures whose validity is to be assessed. For assessing it the writer advocates the use of multivariate analysis and the discriminant function; but he does not explain how these methods would be applied or quote their results. The substance of his review seems excellent; but his style is by no means lucid. If we interpret him correctly, he holds that two main problems call for special study. First, how far does the present system " select or reject the right pupils "—or, as he puts it later, " to what extent do the predicting tests place pupils in the right order of their potentiality to succeed ? " Secondly, what should be " the size of the grammer school group "—or, in other words, " does the system take too many or too few ? "

But there is, as he points out, a prior question to be settled : how are we to judge the success ? The verdicts of the secondary school teacher are, he believes, not sufficiently trustworthy. Nor is he satisfied with an external examination like the ' G.C.E. '. The best criterion, he thinks, would be an internal examination conducted by the school. Personally we are tempted to argue that profiting by what is intended to be a liberal type of education should be judged, not by any sort of examination, but on a genuinely liberal basis. However, the evidence actually available rests, not on a single criterion, but on " the cumulative effect of numerous investigations " : (the more important results are tabulated in a useful appendix); and the outcome " greatly increases our confidence in the efficiency of the selection procedure ". But at this point the Working Party seems to become unexpectedly optimistic. " Throughout this report," says an inserted footnote, " we have assumed that the correlation of an efficiently conducted selection procedure with a perfectly reliable scholastic criterion is close to 0·90." This refers to " validities for over 5 years ". And the practical conclusion drawn is that " little gain in predictive efficiency can be expected from further tinkering "[1].

But is not the figure suggested a little too high ? Since mental growth during adolescence is itself liable to unpredictable fluctuations and since a child's eventual achievements must be influenced by these and many other unforeseen accidents, there can be no such thing as a ' *perfectly* reliable scholastic criterion ' in any literal sense. Moreover, even the reliability coefficient of the selection procedure seldom rises as high as 0·90. The validity coefficients which approach this figure have all been obtained by ' boosting ' the observed value to

[1] In the final chapter on ' Recommendations ' the essential conclusion is more carefully expressed : " the intelligence test is so consistently successful that it should not be dispensed with, *though it might be improved* ".

allow for selection. The observed correlations, we are told, are " typically of the order of 0·45 ". Consequently a theoretical value of 0·90 would virtually double the observed coefficient.

However, having assumed a validity of 0·90 and a selection-ratio of 20 per cent, the writer infers that the number of pupils wrongly placed would be about 10 per cent of the total; one quarter of those selected (5 per cent of the total) would not really deserve their places, while another 5 per cent (1 in 16 of the failures) would presumably have made good in the grammar school had they been given the chance. Now this method of assessing the results implies a further assumption which seems scarcely tenable. Why must we suppose that the number of pupils who genuinely deserve places (as judged by the criterion) must be exactly equal to the number who are given the chance (as judged by the selection procedure) ?

Before we can properly settle this point, we must take up the second of the problems mentioned above; and this in turn involves two distinct questions.

First of all, how many should we *actually* transfer ? The Consultative Committee, it may be remembered, suggested an entry of about 20 per cent, with wide variations for different areas. The Working Party as a whole favours a much higher proportion, namely, 25 to 30 per cent. On the other hand, the only educational adminstrator included among their members suggests a lower figure—15 per cent. The main principle governing the Working Party's argument is apparently an overriding anxiety lest injustice should be done to the less favoured individuals. But in our efforts to give the benefit of the doubt to every questionable case, we may do serious injustice to the community : for we shall almost inevitably lower the standard of teaching and culture in the schools that receive these poorer specimens : and we shall be frittering away our limited financial resources on pupils who at best are rather dubious speculations[1].

But there is another question which we ought to ask : how many are there in the total population who *really can* profit by this more costly type of education ? When a teaching institution in a University selects candidates for training with a view to a degree in medicine, let us say, it does not expect every one of those accepted to pass the final examination : the teaching course itself provides a period of probation. And the number who eventually pass will be much smaller than the number who were originally accepted. Similarly, in accepting candidates for a grammar school, there are two distinct numbers to be borne in mind : (i) we want, so far as we can, to ensure that the ablest members in the total population (the top 2, 5, or 10 per cent or whatever it may be) shall receive an education suited to their ability and aptitudes; but (ii) in order to do so with as few omissions as possible, it will be necessary to admit as many as we can

[1] As many as 18 per cent now leave the grammar school before 16, and 49 per cent at 16 ; and the proportion of early leavers is, as a rule, much greater among pupils from the manual classes and in areas which select a high proportion for entrance : cf. Central Advisory Council for Education, *Early Leaving*, H.M. Stationery Office, 1954. The social and cultural dangers involved in the recent changes in educational organization are ably discussed in *The Uses of Literacy* (1956) by Richard Hoggart, himself a member of the working classes, who profited in this way.

manage, having due regard to the need (*a*) not to waste money on those who are bad bets, (*b*) not to lower the standards of the school, and (*c*) not to impose undue strain on the less able pupils who are accepted. It follows that the total number of transfers should not be equal to, but appreciably larger than, the total number of successes as estimated for the entire population.

The disagreement over the number proposed for transfer by different members of the Working Party seems to reflect a basic weakness in the whole approach. Throughout there is a tendency to assess the practical value of a selection procedure (or of a component test) solely by reference to its theoretical validity. Now what an education committee wants to know is whether the merits of any new procedure will justify the expenditure that it will presumably entail. And this question can hardly be answered unless we can first attempt some fairly definite estimate for both costs and merits.

But, it may be asked, how can this be done except in terms of rough impressionistic arguments ? A careful scrutiny of the literature reveals plenty of parallel examples and a variety of available devices. However, the point is so important, and yet so completely overlooked in the Report, that it seems desirable to cite a concrete illustration. In submitting his budget for expenditure on psychological testing or the like, the Psychologist to the London County Council used regularly to draw up a kind of balance sheet for expected profit and loss, and in doing so assigned to each item a weight depending on its own probability. For instance, in the memorandum on ' The Use of Intelligence Tests and Rating Scales in Examinations for Free Places ', Burt begins with the well known formula for determining ' mathematical expectation ' (given in every schoolbook on higher algebra), viz., ' $x_i = p_i G - q_i L$ ', where i denotes any one of the alternative procedures to be considered, G the gain achieved by selecting a suitable candidate, L the loss entailed from selecting an unsuitable candidate, and p_i and q_i their respective probabilities with this procedure. These expressions are then expanded in terms of their chief components and their respective probabilities, e.g. the cost of the procedure used the amount the Council is willing to spend on educating a suitable candidate at a grammar school, the borderline proposed for a pass, and so on. In summing the ' expectations ' for the several candidates, allowance must be made for the fact that the distinction between pass and fail is relative rather than absolute; and this is effected in the usual way by weighting each candidate according to his marks and weighting each mark by the regression on the criterion. Where two or more alternative policies are to be compared, Burt takes typical combinations of the component variables, calculates the total expectation for each, chooses the minimum, and maximizes that[1].

[1] The use of discrete combinations involves interpolation, which seems a little circuitous. The general principle, it is explained, is based on that used by Edgeworth for his ' hedonic calculus ' (maximizing the minimum values of the expected excess of pleasure over pain : F. Y. Edgeworth, *Mathematical Psychics*, 1910). The method followed would seem in some ways to be the reverse of that which has since figured largely in so called ' acceptance procedures ', viz., minimizing the maximum ' risk '. However, in the case of the hedonic calculus (we are told) " minimizing ' risk ' could most easily be achieved by doing nothing at all—a policy too often adopted in actual practice ".

The conclusions reached afford an excellent instance of the fallacy of basing practical recommendations on theoretical validities alone. The procedure with the highest validity was one which included a separate intelligence test, as well as the usual papers in English and Arithmetic and the teachers' marks; yet it is shown that, in a well-organized educational area like London, the *additional* information supplied by the intelligence test was too slight to warrant the additional cost; while in mixed areas (partly rural and partly urban), it appeared, such a test might amply justify its cost.

In Britain, where the funds at their disposal have been strictly limited, the pioneers of applied psychology, whether in education or in industry, have always paid special attention to questions of profit and cost. And today, even in America, arguments of this kind are now a commonplace. As one educationist has put it, " in every investigation where practical issues are involved, statistical theory should insist that considerations of cost and of the consequences of the decisions should be taken into account "[1]. It is therefore surprising that this aspect of their problem is not touched upon in the Working Party's ' Inquiry '.

A short chapter is devoted to the assessment of attainments in school subjects and to the supplementary techniques available for borderline cases, such as the study of children's interests, the assessment of character qualities, and the value of interviews and oral examinations. An ' oral interview ' (with individual tests) was advocated by the Consultative Committee as a useful adjunct to the ordinary examination; and a recent survey shows that approximately 70 per cent of the various County Councils now adopt some kind of interview for doubtful cases. This the Working Party vigorously condemns. But none of their reasons seems very convincing.

First, we are told, an interview " puts a strain on many children "; and case-studies are quoted to illustrate how a dread of the examination may induce nervous stress. But, as most school doctors would agree, nervous strain is far commoner among the dull but conscientious pupils *after* thay have been transferred to the grammar school than it is among candidates for the examination. A tactful interview would do much to obviate such consequences. Secondly, " interviewers are liable to be impressed by the wrong things, such as deportment or accent ". But it surely should be part of the trained interviewer's technique not only to put the child at his ease, but also to guard against his own personal prejudices. Thirdly, " wartime research confirmed psychologists' doubts as to the consistency of the interviewers and the validity of their judgments ". However, unlike teachers, most wartime psychologists were new to such tasks; and the attempt to judge the personality of adults is a problem quite different from assessing the ability and character of children of eleven.

Finally, it is said, " whenever conclusions derived from interviewing children are allowed to outweigh evidence from test-scores and teachers' estimates, the accuracy of prediction is likely to be lowered ". But the object of the interview

[1] M. A. Girshick (*Rev. Educational Research,* XXIV, 1954, pp. 448f) who gives a useful survey of modern statistical procedures of this type.

is not to secure evidence which shall " outweigh " other estimates, but rather to check and supplement them, particularly where the results of the tests and the estimates of the teacher are ambiguous or at variance. At the same time, it affords a convenient occasion for collecting and coordinating further information about the child's aptitudes, interests, behaviour and past history, for hearing the views of the parents, and above all for making sure that no accident of ill health or transitory disturbance was responsible for the low marks obtained for the written examination. Every grammar school master would prefer to see the borderline candidates himself before any decision is taken as to whether they are suitable for his school, while the head of the primary school would, if present, feel reassured that no case was turned down without full consideration. Certainly the N.U.T. Committee of 1949 looked on the old fashioned interview with disfavour. But their objections scarcely apply to what is sometimes called the ' new style of interview '[1]. In short, the disadvantages alleged seem to be reasons, not for discarding the interview, but for improving it.

The report concludes with thirty-two italicized recommendations. The majority are excellent in substance, though not always felicitously worded. At present, of the local education authorities in England and Wales, barely one quarter enlist the services of a psychologist. " Psychologists," says the report, " should uphold their claim to be able to diagnose the most suitable form of schooling." But the ability to ' diagnose ' a suitable school is primarily an educational rather than a psychological problem; and the majority of educational psychologists are more familiar with the diagnosis of dull, backward, and neurotic pupils than with the ' diagnosis ' of schools. What the psychologist can do, provided he possesses the requisite training and experience, is to help in the construction, standardization, and application of tests, and offer advice on individual children whose ability or character seems problematic. But much of this advice could and should be given long before the stage of selection is reached.

With the more practical recommendations this is scarcely the place to deal. But the detailed suggestions as to the theoretical problems calling urgently for research may well be brought to the notice of our readers. The following seem the most important : " the broader aspects of benefiting from a grammar-school course, and their predictability," " the development of intellectual capacities in adolescence ", " ' old-type ' and other forms of examination or assessment which, without loss in predictive efficiency or undue unreliability in scoring, might have beneficial rather than harmful backwash effects," and finally " the general scaling of school marks and estimates " : (on this last problem there is a most instructive statistical note in one of the appendices). In all these inquiries, the Working Party rightly " insists that the procedures should be planned and supervised by persons with adequate psychological and statistical training ".

To the statistical psychologist perhaps the most interesting feature of the whole inquiry will be its incidental and unanimous vindication of statistical

[1] See Mr. Skeet's account of the ' new style interview ' (as worked out, e.g. by the Northumberland authority in collaboration with Professor Thomson) : *The Child of Eleven*, pp. 131–135.

procedures as the most objective and profitable means available for studying complex practical issues like that of selection. At the same time both its merits and its shortcomings illustrate the supreme importance of combining the theoretical approach of the statistical psychologist with the practical experience of the teacher and administrator. Many of the investigations cited have been carried out by research-students, unfamiliar with the conditions in the classroom and the home, and often working under purely academic guidance. It therefore seems high time that a larger number of education committees should follow the example already set by one or two of the most progressive, and appoint experienced psychologists as members of their own administrative staffs with full authority to undertake systematic surveys and cooperative studies, scientifically planned, of the various problems arising within their areas.

A. G. Houghton
D. R. Morris

BOOK REVIEWS

Statistical Methods and Scientific Inference. By R. A. Fisher. London : Oliver and Boyd, 1956. Pp. viii+175. 16*s*.

This is a book which every statistical psychologist should read. The author explains that in his two previous volumes he had in view two practical aims : in *Statistical Methods for Research Workers* his object was to modify traditional statistical procedures (based on the theory of infinitely large samples) in order to meet the needs of experimental work ; in *The Design of Experiments* the aim was to improve the quality of the data obtained in such work by considering in advance the logical nature of the procedures involved. In these earlier volumes, however, space had not permitted a full discussion of " those refinements of the reasoning process which have come clearly into view with a study of these two aspects of scientific work ". It is with these further refinements that Sir Ronald Fisher's new book is mainly concerned.

Karl Pearson's statistical and biological concepts are somewhat severely criticized. His mathematics are described as clumsy and sometimes misleading (e.g. in problems of distribution); and " in his dispute with Bates on over Mendelian inheritance ", we are told, " he was the bull to the skilful matador ". The newer theories of decision functions are recognized as useful in the fileds of industrial and commercial psychology for which they were originally designed ; but their adoption as aids to forming correct scientific conclusions is firmly rejected.

W. R. Norton

An Introduction to Cybernetics. By W. Ross Ashby. London : Chapman and Hall, Ltd., 1956. Pp. x+295. 36*s*.

Dr. Ashby's aim has been to provide an introductory textbook on ' cybernetics '. His definition of the term is " the science of control and communication in the animal and the machine : in a word, ' steersmanship ' ". As he explains, the present volume is complementary to his earlier work, *Design for Brain*. It falls into three main parts : the first discusses processes developing within a single system; the second, processes of communication between one system and another; and the last, the problem of regulation and control in biological systems. No knowledge of mathematics is required beyond elementary algebra; and each section is followed by graded exercises with hints and answers.

Starting from a few elementary concepts—' operands ' (i.e. states or properties of ' systems '), ' operators ', ' transitions ' (i.e. the changes effected by the operations), repeated transitions, and the like—the author shows how each can be made more precise by formal definition and classification. A complex set of transitions, produced by operating simultaneously on a number of ' operands ' within the same system, is called a ' transformation ', and is represented symbolically by the multiplication of matrices. The series of states exhibited by such a system at successive times is expressed by the successive application of the appropriate transformation, and is said to define a ' trajectory ' or ' line of behaviour '.

In applying this formal procedure to the solution of practical enquiries, the first step is to relate the behaviour of the concrete physical mechanism to the abstract properties of the symbolic expressions—that is, of the matrices and the matrix-operations. We are then introduced to the conception of a ' determinate machine '. This is defined as ' that which behaves in the same way as a closed single-valued transformation '. The definition, it will be noted, turns not on the nature of a material structure, but on its mode of behaviour. The point is illustrated by what the electrical engineer would call ' the problem of the black box ' : we are confronted with an apparatus possessing terminals for input to which he may apply any stimulation he pleases, and terminals for output, which will, enable him to observe the results : the box itself cannot be opened; and the problem is to describe its behaviour or infer its inner structure in the light of the observations so made.

In Part II we are introduced to a mechanism whose behaviour is stochastic, i.e. " determinable only in a statistical sense " ; and the concepts and principles explained in the first part of the book are progressively developed so as to form a basis for the theory of such processes as coding, communicating information, regulating behaviour, feed-back in its various forms, and the attainment of stability or equilibrium. Throughout this type of analysis the more familiar questions of energy play little or no part ; and often the most important variables take the form of probabilities. Moreover, unlike the methods generally adopted in mathematical physics, those proposed by Dr. Ashby are not restricted to metrical characteristics or to processes that are continuous and linear. Since in the biological sciences the behaviour investigated is frequently non-linear and non-continuous, and at times not even metrical, such techniques are naturally of special interest to workers in those fields.

Psychology itself is a biological subject; and a number of psychological problems are briefly discussed by Dr. Ashby to illustrate his points. A special characteristic of the ' mechanisms ' studied by the psychologist is that they seek goals. And goal-seeking is, Dr. Ashby tells us, " an example of behaviour that is stable around a state of equilibrium " ; nevertheless, as he points out, not all the examples of stability so attained are ' goals ', since not all of them are ' desirable '. Here the reference to ' desire ' apparently implies the notion of 'value ' or ' satisfaction ' which, it would seem, can never be reduced to purely mechanistic concepts. Generally, he suggests, " it should be possible for the cybernetically minded psychologist to devise a topological approach which will enable him to get the information he needs ". The book ends with an interesting section on " amplifying the intelligence " (in a rather special sense of that term). Problem-solving, he maintains, is essentially a mater of appropriate selection; and, since power of selection can be amplified, " it seems to follow that intellectual power, like physical power, can be amplified ". As an instance of such amplification he cites the way " gene patterns grow up to form a brain ".

Dr. Ashby's psychological illustrations have been somewhat severely criticized by psychologists themselves; and this has led many to suppose that his methods would be of little use in psychology. The present reviewer would dissent strongly from that verdict. The proposals outlined in his book might perhaps have carried more conviction had he referred in fuller detail to researches already carried out by psychologists, where similar concepts and procedures have already been employed with success. Stout, for examples, over half a century ago asked " whether we can find a mode of physical self-determination, which, like the apperceptive and conative processes studied by the psychologist, is directed towards an end, and ceases when the end is attained "; and found it in the process of ' recovering equilibrium after disturbance ' by means of ' adaptation to irregularly varying outer condition ' through ' indirect ' (or ' re-entrant ') ' self-determination ' (i.e. through feedback). The temporal sequence of states which results he terms a ' vital series ' (an inappropriate phase borrowed from Avenarius) (*Analytic Psychology*, I, pp. 151f.).

Lewin attempted a ' topological ' approach (*Principles of Topological Psychology*, 1936); but at that date the theory of analysis situs was not sufficiently advanced for much progress to be made. Burt's treatment of the subject, familiar to most readers of this *Journal*, comes nearest to that envisaged by Ashby. In his terminology Ashby's ' systems ' become ' persons ', the ' operands ' ' traits ', the ' temporal series ' ' occasions ', and the ' operators ' ' factors ' (i.e. matrix multipliers containing ' weights ' or ' loadings '). If necessary, qualitative traits may be studied instead of metrical by substituting association coefficients for correlation coefficients. Following Stout, Burt regards the essential process to be the recovery of equilibrium after disturbance; but instead of a single stimulus or response we have patterns, expressed by vectors or matrices. The use of matrix models, taken over by Burt from Sheppard, is of long standing in psychology. The application of the ' principal axis transformation ' then enables the investigator to express the whole process in ' canonical ' form; and the use of O-technique, based on time, in place of the more familiar T- or P-techniques renders it possible to " deal with dynamic instead of merely static problems ". Thus the processes of learning and growth are expressed by repeated multiplication by a factor matrix consisting of ' transitional probabilities ', leading, in the simplest case, to a Markov chain (also suggested by Ashby); and what Ashby calls ' trajectories ' are expressed by ' autocatalytic curves '[1]. These detailed similarities between the procedures worked out by two independent investigators seem to offer strong confirmation of their value.

One common criticism advanced by psychologists is that the problems attacked along these lines by statistical investigators (e.g. those of mental inheritance, learning, and the like) cannot be solved until the underlying conditions have been fully elucidated by the experimental geneticist or the neurologist. To this objection the analogy of the engineer studying his ' black box ' seems to afford a convincing reply : after all one can discover a great deal about the behaviour of a new wireless set by twiddling knobs and noting the results without ever peering inside the case of looking at the diagram of the wiring. The psychological problems themselves are of such practical importance that it is urgently desirable to reach at least a provisional solution without waiting half a century for the internal mechanisms to be revealed in detail. For those who are contemplating research in these fields Dr. Ashby's book forms one of the most suggestive contributions published during recent years.

T. F. Ramsay

[1] Cf. this *Journal*, II, pp. 95f., III, pp. 166f., IV, pp. 179f., and IX, pp. 49f. For some of the analogies between factorial methods and cybernetic methods I am indebted to a lecture given by Professor Burt at Worcester.

INDEX TO VOLUME X

INDEX OF AUTHORS

	PAGE
Bernyer, G. : Psychological Factors : their Number, Nature, and Identification	17
Burt, C., and Howard, M. : Heredity and Intelligence : A Reply to Criticisms	33
Burt, C., and Howard, M. : The Relative Influence of Heredity and Environment on Assessments of Intelligence	99
Dunsdon, M. I., and Roberts, J. A. F. : A Study of the Performance of 2,000 Children of Four Vocabulary Tests	1
Hotelling, H. : The Relations of the Newer Multivariate Statistical Methods to Factor Analysis	69
Howard, M., and Burt, C. : Heredity and Intelligence : A Reply to Criticisms	33
Howard, M., and Burt, C. : The Relative Influence of Heredity and Environment on Assessments of Intelligence	99
Lewis, W. B. : Professor Lewis M. Terman	65
Roberts, J. A. F., and Dunsdon, M. I. : A Study of the Performance of 2,000 Children on Four Vocabulary Tests	1
Stuart, A. : The Comparison of Frequencies in Matched Samples	29
Walker, D. A. : The Theory and Use of the Success-Ratio	105
Wrigley, C. F. : The Distinction between Common and Specific Variance in Factor Theory	81

INDEX OF SUBJECTS

Common and Specific Variance in Factor Theory, The Distinction between : Wrigley, C. F.	81
Frequencies in Matched Samples, The Comparison of : Stuart, A.	29
Heredity and Intelligence : A Reply to Criticisms : Burt, C., and Howard, M.	33
Heredity and Environment on Assessments of Intelligence, The Relative Influence of : Burt, C., and Howard, M.	99
New Multivariate Statistical Methods to Factor Analysis, The Relations of the : Hotelling, H.	69
Psychological Factors : their Number, Nature, and Identification : Bernyer, G.	17
Success-Ratio, The Theory and Use of the : Walker, D. A.	105
Terman, Professor Lewis M. : Lewis, W. B.	65
Vocabulary Tests, A Study of the Performance of 2,000 Children on Four : Dunsdon, M. I., and Roberts, J. A. F.	1

BOOKS REVIEWED

Ashby, W. Ross : *An Introduction to Cybernetics*	126
Cronbach, L. J., and Gleser, G. C. : *Psychological Tests and Personnel Decisions*	64
Fisher, R. A. : *Statistical Methods and Scientific Inference*	125
Sholl, D. A. : *The Organization of the Cerebral Cortex*	64
Vernon, P. E. (ed.) : *Secondary School Selection*	113

BIBL. LONDIN. UNIV.

PRL92

Vol. X Part II November 1957

TP1

THE BRITISH JOURNAL OF STATISTICAL PSYCHOLOGY

EDITED BY
CYRIL BURT

WITH THE ASSISTANCE OF
CHARLOTTE BANKS AND ALAN STUART
AND THE FOLLOWING EDITORIAL BOARD

A. C. AITKEN
M. S. BARTLETT
W. G. EMMETT
M. G. KENDALL
D. N. LAWLEY
E. S. PEARSON
E. A. PEEL
L. S. PENROSE
J. FRASER ROBERTS
A. RODGER
W. STEPHENSON
P. E. VERNON

Managing Sub-editor J. W. WHITFIELD

Printed and Published by
TAYLOR & FRANCIS LTD.
RED LION COURT, FLEET STREET, LONDON, E.C.4

Price 20*s. per part* (*U.S.A.* $3.50) *Subscription* 30*s.* 6*d. per volume* (*U.S.A* $4.50)

PUBLICATIONS OF THE BRITISH PSYCHOLOGICAL SOCIETY

The British Journal of Statistical Psychology is issued by the British Psychological Society. The subscription price per volume is 30*s*. net for Members of the Society and 30*s*. 6*d*. (post free) for non-members. The subscription price in the U.S.A. is $4,50 (post free). Members of the Society should send their subscriptions to THE SECRETARY, British Psychological Society, Tavistock House South, Tavistock Square, London, W.C.1 : non-members to Messrs. TAYLOR & FRANCIS, LTD., 18 Red Lion Court, Fleet Street, London, E.C.4. Until further notice the Journal will be issued in two parts each year, in May and November.

Papers for publication should be sent to Sir CYRIL BURT, 9 Elsworthy Road, London, N.W.3. Authors will receive 25 copies of their articles free; extra copies may be purchased provided the order is given when the proofs are returned. Contributors should indicate whether they wish their reprints to be bound in covers. The printers will state current prices when sending out proofs.

Books for review should be sent to the Editor, advertisements to the Secretary, British Psychological Society.

The Society also issues quarterly *The British Journal of Psychology* and *The British Journal of Medical Psychology* : subscriptions, 60*s*. per annum for either journal, should be sent to the Cambridge University Press, Bentley House, Euston Road, London, N.W.1. The Society publishes, jointly with the Association of Teachers in Colleges and Departments of Education, *The British Journal of Educational Psychology* : for this the subscription, £1 per annum, should be sent to Methuen & Co., Ltd., 36 Essex Street, London, W.C.2. Members of the Society receive the foregoing journals on special terms; enquiries should be addressed to the Secretary, British Psychological Society.

Papers for publication in *The British Journal of Psychology* should be sent to Professor James Drever, Department of Psychology, the University, Edinburgh; those for publication in *The British Journal of Medical Psychology* to Dr. Joseph Sandler, 96 Portland Place, London, W.1; those for publication in *The British Journal of Educational Psychology* to Professor P. E. Vernon, London University Institute of Education, Malet Street, London, W.C.1.

PREPARATION OF PAPERS

Contributors are asked to send their papers in a form which is ready for submission to the printer : that is to say, the arrangement of the material should follow that observed in previous publications of this *Journal*. Each article should be headed either with a series of numbered headings in italic, corresponding with the cross-headings of the several sections, or with a brief abstract : (the latter procedure is preferred). Each section should have a short cross-heading, in capitals; and the whole should conclude with a summary embodying the main conclusions reached. Special attention should be paid to such details as correct spelling, grammar, capitalization, numbering, etc. (particularly in the case of tables and references). In preparing manuscripts and in correcting proofs authors should conform, so far as possible, with the ' Recommendations to Authors ' set out in *The Printing of Mathematics* by T. W. Chaundy, P. R. Barrett, and Charles Batey (Oxford University Press, 1954, ch. II, pp. 21–73), or to those given in the paper on ' Setting Mathematics ' (Arthur Phillips, *Monotype Recorder*, XL, iv, 1956).

Owing to the present conditions for printing, the cost of setting up tables and algebraic formulae which require hand composition is very much greater than the cost of machine composition. Contributors are therefore advised to arrange their material so that it can be printed as economically as possible. Often numerical results and algebraic formulae can, with a little simplification, be readily adapted for purposes of machine composition. Manuscripts involving many tables and numerous equations can as a rule only be accepted if the authors are willing to pay for the additional cost.

In future authors will receive two complete slip-proofs of their papers, but no page proof. They will be expected to pay the cost of all corrections for which they themselves are responsible.